

# **Nutrition, Physical Activity and Obesity Data Analysis**

IE6600 – Computation and Visualization for Analytics

Final Report

**Group Number 02**

Sparsh Marwah (002246631)  
Sharvari Deshpande (002293025)  
Param Madan (002278621)  
Rukmini Reddy (002830088)

## **Introduction**

In the US, obesity, physical inactivity, and nutrition are the top three public health issues. Sedentary lifestyles, poor dietary choices, and high rates of obesity are a few chronic health conditions that exacerbate diabetes, heart disease, and some types of cancer. To address these issues, campaigns to encourage regular exercise and better eating habits are being launched in addition to policies that support healthier lifestyles. The prevalence of obesity is still rising despite these efforts, which emphasizes the need for all-encompassing, long-term strategies to stop this epidemic and enhance public health.

The Nutrition, Physical Activity, and Obesity dataset offers a thorough and in-depth understanding of adult health behaviors and outcomes in the US across various demographic groups. This dataset is a useful tool for comprehending the intricate relationship between lifestyle decisions and health outcomes because it includes a wide range of variables related to nutrition, physical activity, weight status, and sociodemographic characteristics.

The purpose of this report is to summarize the abundance of data in the dataset in order to pinpoint important patterns, trends, and discrepancies pertaining to physical activity, obesity, and nutrition. We can develop targeted interventions and policy initiatives by gaining a nuanced understanding of health behaviors and outcomes through the examination of data at both the national and state levels.

This report will delve into trends in obesity, nutrition, and physical activity, alongside the development of dynamic dashboards using web frameworks. These dashboards will facilitate interactive exploration of demographic factors within the dataset, aiding stakeholders in identifying insights for tailored interventions.

## **Data Sources**

The dataset that has been used in this project has been acquired from the [dta.gov](https://data.dhs.gov/) website. The dataset encompasses a comprehensive collection of information derived from the Behavioral Risk Factor Surveillance System (BRFSS), focusing on adult dietary patterns, physical activity levels, and weight status. Originating from the Division of Nutrition, Physical Activity, and Obesity (DNPAO), this dataset serves as a crucial component of the DNPAO's Data, Trends, and Maps database.

This database is a robust resource offering both national and state-specific data on key public health indicators related to obesity, nutrition, physical activity, and breastfeeding practices. It provides a nuanced understanding of health behaviors and outcomes across different demographic groups and geographic regions within the United States.

Demographic factors crucial for understanding disparities and trends in health outcomes are also included in the dataset. These factors comprise age, education, gender, income, and race/ethnicity, providing insight into how socio-economic and cultural variables intersect with health behaviors. Additionally, geolocation information enhances the dataset's utility by enabling spatial analysis and visualization of trends across different states and regions.

## **Data Cleaning**

The data cleaning process involved several steps to ensure the dataset's integrity and usability for analysis:

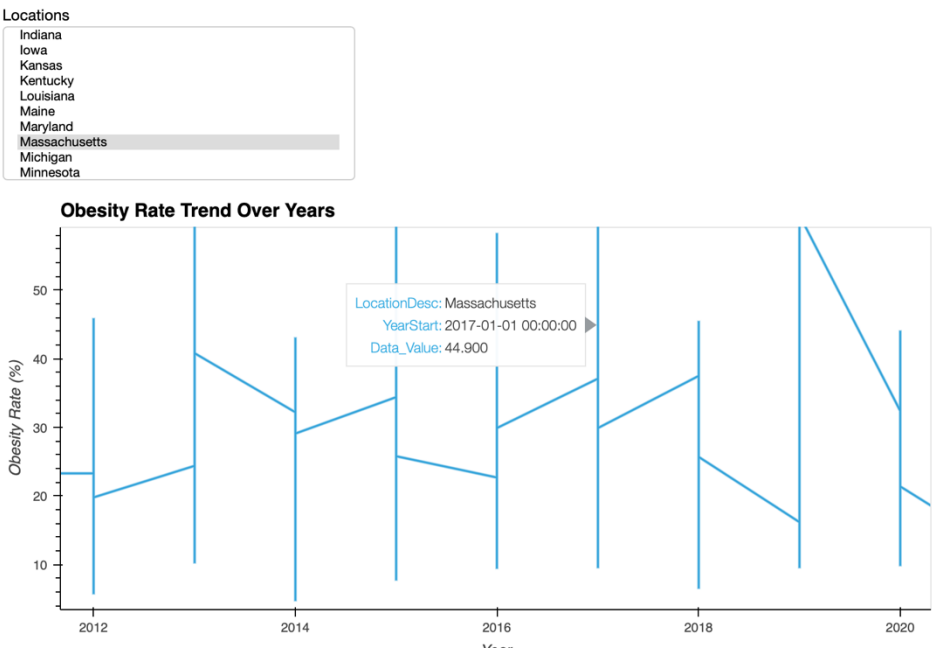
- Initial Assessment: The structure of the dataset, including the number of rows and columns, data types, and missing values, was first determined by looking at it.
- Handling Missing values: 'Data\_Value', 'Age(years)', 'Education', 'Income', 'Race/Ethnicity', and 'GeoLocation' were among the several columns where missing values were found. In order to fix this, missing values were removed from the 'Data\_Value' column, which is essential for analysis. To maintain the distribution of the data, missing values for categorical variables such as "Income," "Race/Ethnicity," "Age(years)," and "Education" were imputed with the mode, or most frequent value.
- Dropping Unnecessary Columns: The dataset was purged of columns like "Data\_Value\_Unit," "Data\_Value\_Footnote\_Symbol," and "Data\_Value\_Footnote" because it was determined that they contained redundant or null values.
- Managing Category Variables: Geographic information was included in the 'GeoLocation' column.
- Final Review and Summary: After completing the cleaning process, a final review of the dataset was conducted to ensure that all necessary steps had been taken. The cleaned dataset consisted of 84,014 entries and 28 columns, containing essential variables such as 'YearStart', 'YearEnd', 'LocationAbbr', 'Datasource', 'Class', 'Topic', 'Question', 'Data\_Value\_Type', 'Data\_Value', 'Low\_Confidence\_Limit', 'High\_Confidence\_Limit', 'Sample\_Size', 'Age(years)', 'Education', 'Income', 'Race/Ethnicity', 'GeoLocation', 'ClassID', 'TopicID', 'QuestionID', 'DataValueTypeID', 'LocationID', 'StratificationCategory1', 'Stratification1', 'StratificationCategoryID1', and 'StratificationID1'.

## Analysis and Visualization

The data has been analyzed to understand the trends This dashboard offers an interactive platform for exploring nutrition, physical activity, and obesity rates across different demographic groups and years. By leveraging various widgets, users can filter and manipulate data views dynamically, enabling them to uncover insights into diverse trends within the dataset.

Out[21]:

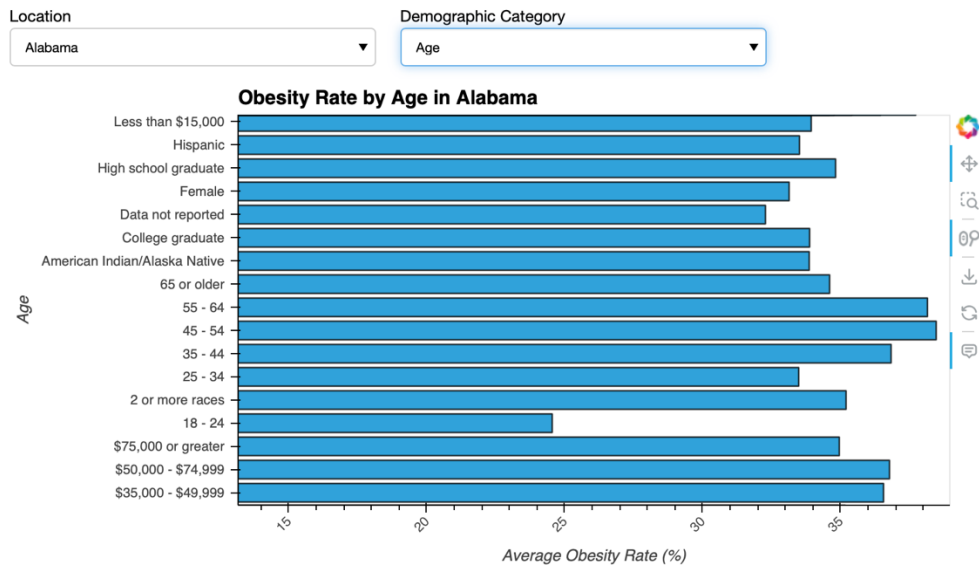
### Obesity Rate Trend Over Years



This visualization showcases the obesity rates across different states over a span of 10 years. Users can dynamically select a specific state from a dropdown menu to focus on its obesity trends. This interactive feature allows users to explore how obesity prevalence varies from state to state and track changes over time within their chosen state. When the users place the cursor on the graph it displays Location, date and date value attributes.

Out[26]:

## Demographic Breakdown of Obesity Rates



Out[38]:

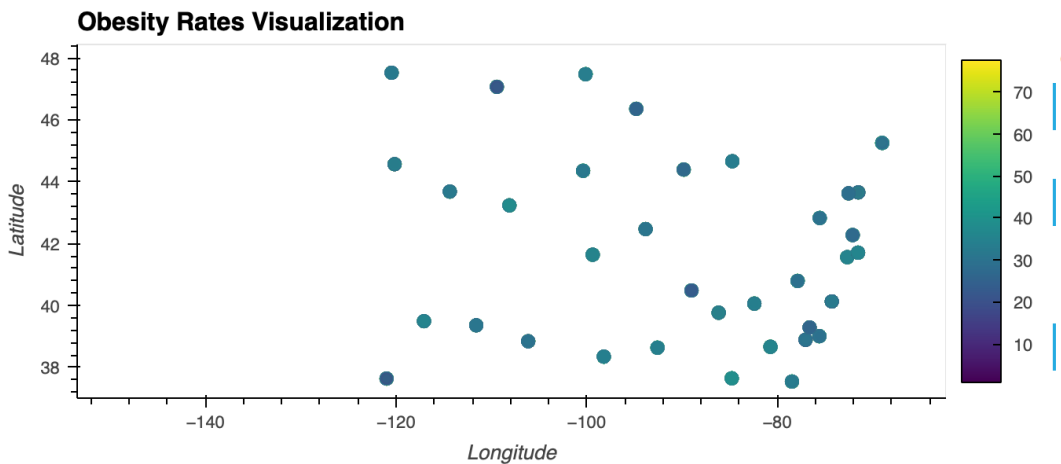
## Demographic Breakdown of Obesity Rates



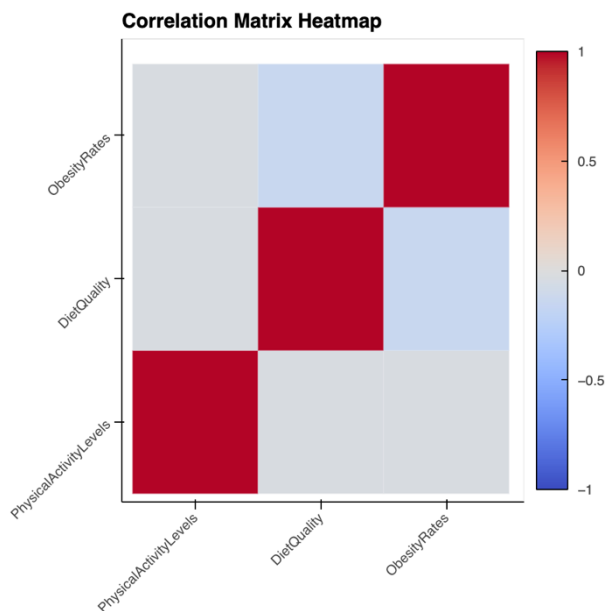
The above visualization shows the average obesity rate with respect to all the available demographics, the demographic category drop down has the options age, education, income and race /ethnicity. We can change the location and demographic category, this will give us a different visualization.

The demographics on the Y axis are taken from the dataset they are a wide range of demographics and we get to select a particular demographics category and see the relation of that with all the others on the Y axis.

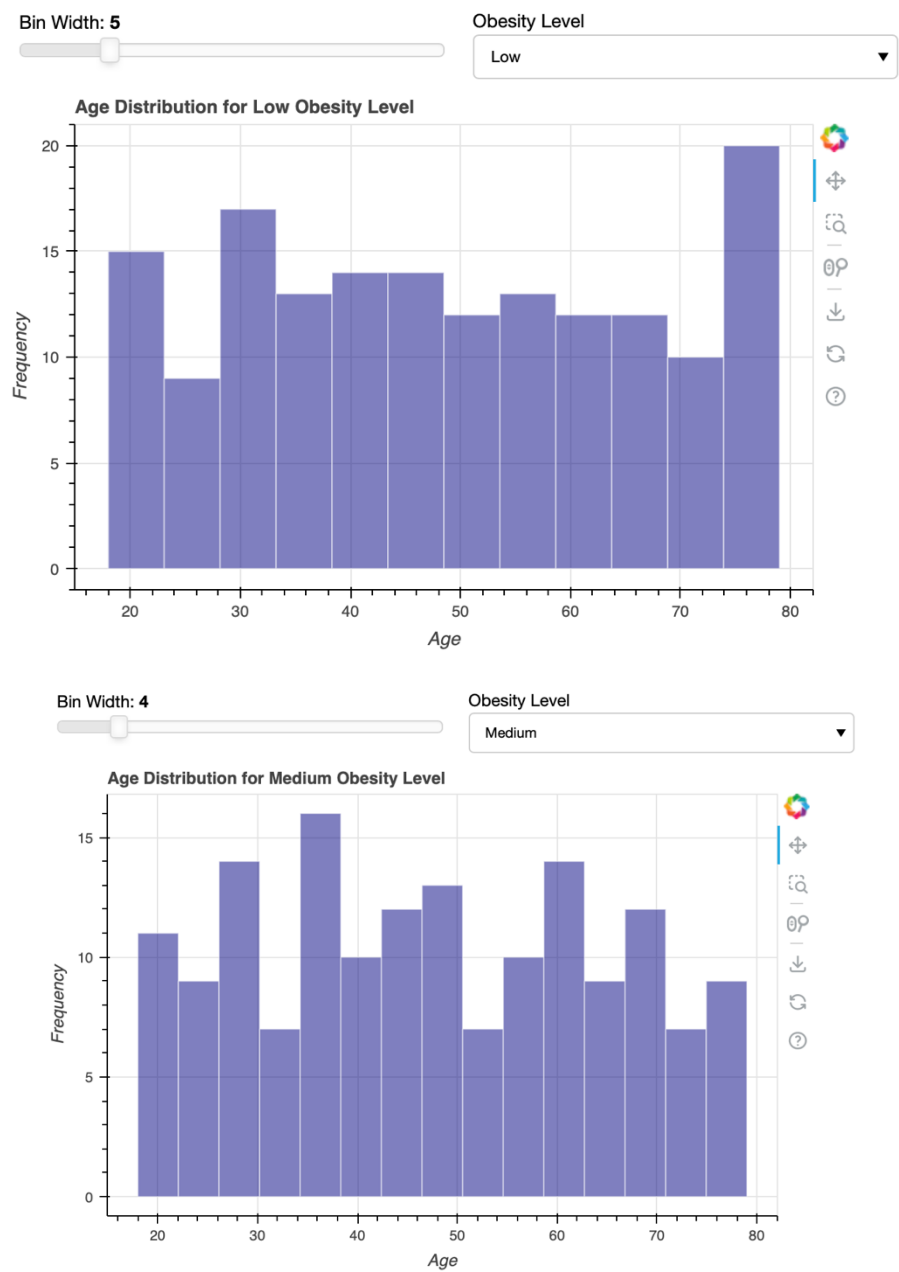
## Interactive Obesity Rates Visualization



The visualization presented above offers insights into obesity rates across diverse states, employing dots whose colors vary to reflect the degree of obesity within each state. A scale positioned adjacent to the visualization aids in understanding the correlation between color and obesity rate. By hovering over individual dots, users access to detailed information such as the name of the respective state, enriching their understanding of the data.

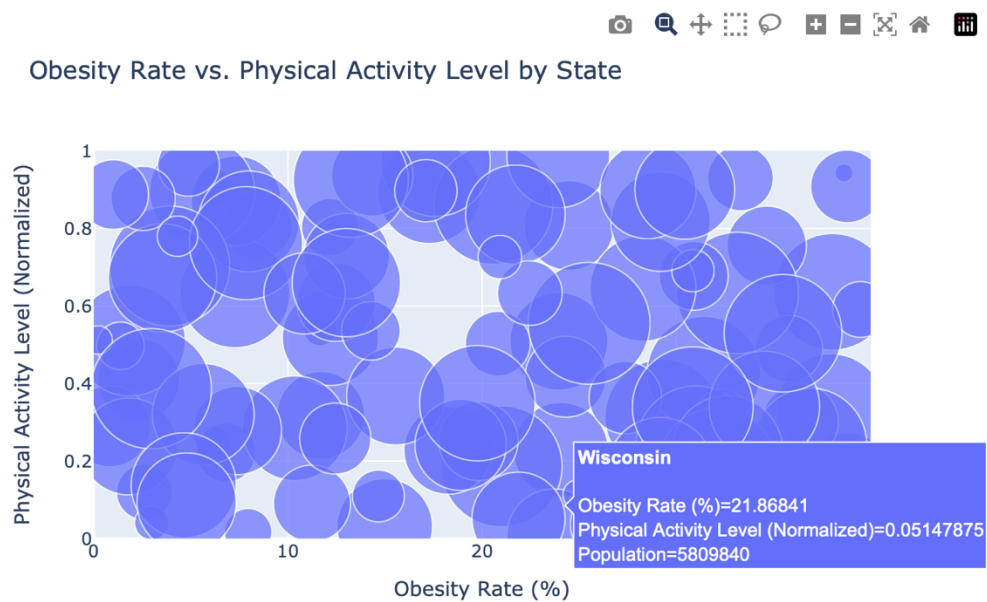


The relationships between various variables, such as physical activity levels, diet, and obesity rates, are visually represented by the correlation matrix heatmap. Notably, it shows that diet quality and obesity rates have the strongest correlation, providing insight into the possible impact of dietary practices on the prevalence of obesity. This realization implies that additional research into the connection between obesity and diet may be especially helpful in recognizing and resolving public health issues pertaining to weight control.





This visualization depicts the frequency distribution of individuals with medium and low-level obesity across different age groups. The data categorizes obesity levels into three categories: low, medium, and high. Additionally, it offers the flexibility to adjust the bin width according to specific requirements, allowing for customization of the age groupings displayed on the graph.



This visualization represents each state as a bubble, with the size of the bubble corresponding to the population of that state. As the user hovers the cursor over a specific bubble, they can view the state's name, obesity rate, physical activity level, and population.

## **Conclusion**

In conclusion, examining the Nutrition, Physical Activity, and Obesity dataset has shed light on the intricate relationship between lifestyle decisions and health outcomes in the US. Through the comprehensive analysis of a wide range of variables, including weight status, physical activity, nutrition, and demographics, we have developed a thorough understanding of public health trends and disparities at the state and national levels.

This analysis has benefited greatly from the use of dynamic dashboards and interactive visualizations, which provide clear means of understanding complicated data and spotting important patterns and trends. With the use of these tools, stakeholders are now better equipped to develop evidence-based strategies for encouraging healthier lifestyles and lowering the burden of chronic diseases as well as to effectively navigate the complexities of public health data.

Users can use these dynamic dashboards to explore demographic factors in order to identify disparities and provide information for focused policy initiatives and interventions. The correlation matrix heatmap has provided insightful relationships between variables, pointing out areas that require targeted intervention and highlighting the critical role that dietary practices play in addressing the obesity pandemic.

Furthermore, research on obesity rates over time and among different demographic groups in different states has shed light on regional differences and variations. These understandings are essential for directing interventions and allocating resources to the most underserved areas.

All things considered, the application of dynamic dashboards and interactive visualizations has proven crucial in revealing patterns, assisting in decision-making, and directing actions to tackle the complex issues brought about by obesity, physical inactivity, and inadequate dietary habits in the US. Stakeholders can create evidence-based plans for encouraging healthier lifestyles and enhancing public health outcomes across the country by utilizing these tools.