Introduction

For this project, I chose the "Nutrition, Physical Activity, and Obesity – Behavioral Risk Factor Surveillance System" dataset from data.gov, spanning from 2011 to 2021 [1]. This dataset offers extensive information on adult diet, physical activity, and weight status, essential for the DNPAO's Data, Trends, and Maps database. The original dataset contained over 88,000 records.

In my analysis, I sought to answer 4 questions:

1) What is the relationship between income and physical activity?
2) What is the relationship between education and obesity rates?
3) What is the relationship between education and physical activity?
4) What is the relationship between income and obesity rates?

About the Dataset

The dataset is 33.2 megabytes in size [1]. This dataset includes data on adult's diet, physical activity, and weight status. It is a collected dataset of multiple studies on the topics of nutrition, exercise, income, education level, age, gender, and other categories. The metadata includes information about various data items, such as 'YearStart,' 'Topic,' 'LocationAbbr', and 'Question' and various data values. The dataset contains various types of nominal, interval, ratio, and ordinal data. Data is both numeric, temporal, and plain text.

This data was collected by the Behavioral Risk Factor Surveillance System in the US Department of Health and Human Services (HHS) and published by the Center for Disease Control and Prevention (CDC) [1]. According to the HHS website, the departments mission is to "enhance the health and well-being of all Americans, by providing for effective health and human services and by fostering sound, sustained advances in the sciences underlying medicine, public health, and social services." [2] The CDC oversees detecting and responding to national health threats, tackling the biggest nationwide health issues, advancing technology and science to prevent disease, and promoting individual and community healthy behaviors [3].

This data is used for the Division of Nutrition, Physical Activity, Obesity (DNPAO) within the CDC to track data and trends and provide national and state data on obesity, nutrition, physical activity, and breastfeeding to get a pulse of the nation's health and direct policy towards improving national health metrics [1]. This is a big data problem because the dataset is a made up of many studies and surveys on the topics of health. The dataset is quite robust in the variance of metrics. The variance in metrics makes this dataset particularly challenging to break down to draw usable conclusions and inferences due to its variety in data types, incomplete categories, and formats.

The CDC privacy and ethics policies are quite restrictive to protect research subject's private information. The CDC website does not collect any personal identifiable information while visiting their website [4]. The CDC also abides by the Freedom of Information Act which requires them to release all findings to the public and to provide information to the public upon request [5]. Lastly, the CDC has a Vulnerability Disclosure Policy which protects the American public's information from unwarranted disclosure [6]. This policy gives clear guidelines to security researchers on conducting vulnerability discovery activities.

Analysis

I created separate statistics and visualizations for each of these questions. The data was extracted, cleaned, and parsed from the original CSV file downloaded from the data.gov website. I found the best way to group the data was by the research question and stratification categories. The research questions were divided into two groups: questions regarding obesity and questions regarding physical activity. From those question groups, only data where the stratification category was either "Income" or "Education" was extracted to be analyzed. Post-cleaning, the number of records did not total more than 10000 records per research question with the lowest being 5000 records extracted that were applicable to the research question "What is the relationship between education and obesity?" The reason for the low number of usable records was due to large quantities of missing relevant information per record. Records with missing relevant information were discarded from the data.

Summary statistics were calculated for each separate research question in the dataset and the statistics were divided up by income brackets and levels of education. Bar charts were created for each question in each stratification category to display the percentages of those in the research study. Bar charts were created for each research question. Due to the structure of the data, bar charts or pie charts were the two best options to visualize the data due to the separation of categories.

The data will be interpreted by analyzing the output for each question.

What is the relationship between income and physical activity?

Individuals with an income of $75,000 or greater per year were more physically active compared to lower income groups. If we look at all the bar charts, we can see that those with an income bracket of $75,000 or greater were significantly more active than other income groups. Individuals making $15000 or less per year had the lowest rates of physical activity. In only one category did individuals who exercise exceed 50% on average. The American Heart Association recommends 150 per week of moderate intensity exercise [7]. The summary statistics in the dataset showed that 'Percent of adults who achieve at least 150 minutes a week of moderate-intensity aerobic physical activity or 75 minutes a week of vigorous-intensity aerobic activity (or an equivalent combination)' only met or exceeded 50% for those making $35,000 a year or more.

What is the relationship between education and physical activity?

Individuals who were college graduates were more physically active compared to lower levels of education. Individuals with less than a high school education were the least active. The summary statistics showed that 'Percent of adults who achieve at least 150 minutes a week of moderate-intensity aerobic physical activity or 75 minutes a week of vigorous-intensity aerobic activity (or an equivalent combination)' were 59% in those who were college graduates compared to 38% for those with less than a high school education.

What is the relationship between education and obesity rates?

There was a less of a difference between education groups when looking at obesity rates. There were two questions in the dataset regarding obesity: 'Percent of adults aged 18 years and older who have obesity' and 'Percent of adults aged 18 years and older who have an overweight classification'. In the overweight category, college graduates had the highest rate of obesity at 37% compared to less than high school at 33.5%. In the obesity category, the opposite is observed. College graduates have the lowest rate of obesity at about 25% while those who have less than a high school education was almost 10% higher at 34%.

What is the relationship between income and obesity rates?

Like the relationship between education and obesity rates, there was less of a difference between groups when divided by income. The summary statistics and graphs tell a similar story with about 38% at the highest income bracket reporting being overweight compared to 29% in the lowest income bracket. In the obesity category, the statistics are reversed with 35.5% of people in the lowest income bracket reporting obesity and 28% in the highest income bracket who are obese.

Conclusion

It appears that income and level of education could have some sort of impact on how much an individual exercises and obesity rates. The higher the income bracket and education level, the less likely the individual is overweight/obese and the more likely they exercise. The lower the income bracket and education level, the more likely the individual is overweight/obese and the less likely they are to be getting enough exercise. While income level and education does seem to influence certain indicators of health, it cannot be concluded that those are the sole determinants. Time constraints, job demands, and lifestyle choice could also all play a role in individuals being unable to achieve the recommended activity level. Even among higher earners, only a marginal majority meet the physical activity guidelines. Factors such as time availability, job nature, environmental aspects, accessibility to exercise modalities, and educational background might significantly influence physical activity levels. Similarly. Education and income level are the sole determinants of obesity. Differences in food accessibility, cultural influences, and environment could also have an impact on the obesity rates across income and education levels.

For future analysis, I would recommend exploring the other factors mentioned above like time constraints and accessibility to food to gain a better understanding of what could be impacting obesity rates and physical activity in the United States. I would also recommend further analysis for why overweight rates are higher in higher income brackets and education levels, yet obesity rates are higher in lower income brackets and education levels. What factors play a role in someone's classification changing from overweight to obese? Overall, further data would be need to collected and analyzed to better understand the United States obesity rates and levels of physical activity.

References

1. Nutrition, Physical Activity, and Obesity - Behavioral Risk Factor Surveillance System catalog. (2023, December 8). https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-behavioral-risk-factor-surveillance-system

2. Assistant Secretary for Public Affairs (ASPA). (2023, October 2). About HHS. HHS.gov. https://www.hhs.gov/about/index.html

3. Mission, role and pledge. (2022, April 29). Centers for Disease Control and Prevention. https://www.cdc.gov/about/organization/mission.htm

4. Using CDC.gov. (2022, November 2). Centers for Disease Control and Prevention. https://www.cdc.gov/other/privacy.html

5. Freedom of Information Act Importance to CDC Mission | CDC. (n.d.). https://www.cdc.gov/od/foia/

6. Officer, O. O. T. C. I. (2023, March 21). Vulnerability Disclosure Policy. HHS.gov. https://www.hhs.gov/vulnerability-disclosure-policy/index.html

7. American Heart Association recommendations for physical activity in adults and kids. (2024, January 19). www.heart.org. https://www.heart.org/en/healthy-living/fitness/fitness-basics/aha-recs-for-physical-activity-in-adults