**Information Seeking Assignment**

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

**1. An APA-formatted data citation**

Nutrition, Physical Activity, and Obesity - Behavioral Risk Factor Surveillance System. (2017, July 14). Retrieved September 10, 2017, from https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-behavioral-risk-factor-surveillance-system

**2. Details of the license or terms of use**

Open Database License (ODbL): It is free to copy, distribute, use the database, produce works from the database, modify, transform and build upon the database as long as it is attributed to the license of the database clearly and any notices on the original database are kept, it is shared alike under ODbL and kept open. The details of the license and terms of used can be accessed form the link below: <http://opendefinition.org/licenses/odc-odbl/>

**3. About one paragraph describing why these data are interesting**

Health expenditures constitute an important part of the government budget through programs such as Medicaid and Medicare which tend to increase with obesity. Private businesses also have difficulty remaining competitive if health expenditures, health related absentia, and health insurance premia increase. This dataset includes national and state specific data on adults aged 18 years or older who have obesity and have an overweight classification as a weight status, their diet data such as percentage of adults who report consuming fruits or vegetables less than once a day, their physical activity data such as percentage of adults who engage in muscle-strengthening activities on two or more days a week and a few of other physical activity indicators as taken from Behavioral Risk Factor Surveillance System. I have found the dataset interesting since the nutrition, physical activity, and obesity can be linked to education level, income level, and race. As can be seen from the dataset, as the education level increases, obesity decreases but I would like to examine and analyze this dataset in depth to find more about the causation/correlation relationship between these variables. Also, this dataset is not that clear to see all the variables and for some of them the data is not available because of the insufficient sample size. I would like to learn more about data-cleaning processes, and I think working with this dataset would give the opportunity for the hands-on experience.

**4. Potential data users and decision-makers for this data**

This dataset can be used by health government agencies such as U.S. Department of Health and Human Services, National Institute of Health, Food and Drug Administration. According to the website that I have retrieved the data from, Center for Diseases Control and Prevention’s (CDC) Division of Nutrition, Physical Activity, and Obesity (DNPAO) used this database since it provides national and state specific data on obesity, nutrition, and physical activity. According to CDC, obesity affects 1 in 3 adults putting people at risk for some chronic diseases which costs US health care system billions of dollars. Since defining the problem is the initial step to prevent it and improve public health, this dataset was used by DNPAO in their effort to protect the health of the nation.

**5. Three questions this data might help to answer; note additional sources needed if applicable**

The three questions that this data might help to answer would be:

1. The relationship between obesity and education. Does obesity decrease as education level increases? The dataset divided the education status as less than high school, high school graduate, some college or technical school and as college graduate.
2. The relationship between obesity and income. Does obesity decrease as income level increases? The dataset divided the income levels as: Less than $15,000; $15,000 - $24,999; $25,000 - $34,999; $35,000 - $49,999; $50,000 - $74,999 and $75,000 or greater.
3. The relationship between obesity and race. The dataset classified races as: Non-Hispanic White, Non-Hispanic Black, Hispanic, Asian, Hawaiian/Pacific Islander, American Indian/Alaska Native, two or more races and other. All of the above variables can be integrated and questions such as “what is the relationship between obesity and income for female Hispanics that are college graduates?” can be analyzed.

**Centers for Disease Control and Prevention (CDC) 1991-2015 High School Youth Risk Behavior Survey Data**

**1. An APA-formatted data citation**

High School YRBS. (n.d.). Retrieved September 10, 2017, from http://nccd.cdc.gov/youthonline/

**2. Details of the license or terms of use**

CDC website states that “it is a public domain Web site, which means you may link to CDC.gov at no cost and without specific permission” and “The CDC/ATSDR FOIA Office follows the rules set forth in the Department of Health and Human Services’ FOIA regulations at 45 CFR PART 5.” The details of the policies can be accessed from https://www.cdc.gov/Other/policies.html.

**3. About one paragraph describing why these data are interesting**

Youth of a country represents its future. If they are healthy with good judgement, then we have a bright future. Otherwise we need to invest more in education and training. I have accessed the data from the Youth Risk Behavior Surveillance System (YRBSS), which I did not know about before. The system monitors six types of health-risk behaviors that contribute to the leading causes of death and disability among youth and adults. These six types include; behaviors that contribute to unintentional injuries and violence, sexual behaviors related to unintended pregnancy and sexually transmitted diseases, alcohol and other drug use, tobacco use, unhealthy dietary behaviors and inadequate physical activity. I wanted to look at Youth Risk Behavior Surveillance System (YRBSS) data from 1991 – 2015 at the national level. I have included all the questions that were asked to high school students so that I can sort and analyze the data on the basis of race/ethnicity, sex, grade, and sexual orientation.

**4. Potential data users and decision-makers for this data**

The potential users for this data could be educators, parents, policy makers, government agencies such as U.S. Department of Education, U.S. Department of Health and Human Services, Food and Drug Administration, the Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF).

**5. Three questions this data might help to answer; note additional sources needed if applicable**

1. 116 questions were asked to high school students related to these six types of risky behaviors. I would like to see the relationship between the time spent watching television and sleep patterns. One of the questions was “watched 3 or more hours per day of television (on an average school day)” which can affect the minimum suggested sleep time of 8 hours for youth that is asked as “Did not have 8 or more hours of sleep (on an average school night)”. What is the relationship between time spent playing video games and having sleep deprivation and is our youth getting less and less sleep each year?

2. Do video games inspire violent behavior? One of the questions asked to high school students was “Played video or computer games or used a computer for 3 or more hours per day (for something that was not school work on an average school day)” and the other question that I want to compare the data with is “Carried a weapon (such as, a gun, knife, or club, on at least 1 day during the 30 days before the survey)”.

3. Is cyber-bullying substitute for physical bullying on school property? The answers to the survey questions that I would like to look at are “Were electronically bullied (counting being bullied through e-mail, chat rooms, instant messaging, Web sites, or texting during the 12 months before the survey)” and “Were bullied on school property (during the 12 months before the survey)”.

**1. An APA-formatted data citation**

O. (n.d.). OECD Statistics. Retrieved September 12, 2017, from <http://stats.oecd.org/>

I have used two datasets and for the first one I have selected from the side menu in this order: Health, Health Status, Perceived Health Status by Age and Gender, and then Perceived Health Status by Socioeconomic Status

For the second dataset, I have selected Productivity, Productivity and ULC – Annual, Total Economy, Growth in GDP per capita, Productivity and ULC.

**2. Details of the license or terms of use**

This is an open to public data and the details can be found at <http://www.oecd.org/termsandconditions/>.

**3. About one paragraph describing why these data are interesting**

The relationship between perceived health status and GDP growth is important because if there is a negative relationship as suggested by Ruhm (2000, 2003) then pro-growth policies may not be as beneficial as initially thought. If there is a gender or demographic difference in this effect, then policies can be developed accordingly. If perceived health status is not going better in time consistently, then the reasons for this is worth exploring, such as problems caused by urbanization.

References:

Ruhm, C.J. 2000. Are recessions good for your health? Quarterly Journal of Economics 115 (2), 617-650.

Ruhm, C.J. 2003. Good Times Make You Sick. Journal of Health Economics 22 (4), 637-658.

**4. Potential data users and decision-makers for this data**

Policy makers, government agencies, economists, health insurance companies.

**5. Three questions this data might help to answer; note additional sources needed if applicable**

1) How are GDP growth and perceived health status related for different demographic groups?

2) How does this relationship change between more industrialized countries and less?

3) Is perceived health status getting consistently better in time for any demography?

The word count is 1496.