

Draft title for CSC 595 Report

Authors:

CSC 595

...

0.1 Introduction and Data Description

Data Introduction and Overall Summary

The NHANES is a large scale longitudinal database created by the Centers for Disease Control and Prevention (CDC) to collect data from a sub-sampling of the United States population. The data includes interview data, physical examinations to include vital signs (systolic and diastolic blood pressure, height, weight, BMI, ...) as dental information, laboratory measurements (i.e. glucose or HBA1C values for diabetes), and demographics information. The purpose of this data is set national standards (i.e. BMI percentile measurements standardized by age and sex), track health data for diseases like cardiovascular disease or diabetes to shape public policy, or provide data for international research organizations and academic institutions for many purposes. Please see the following example of areas which have used this data: Iranpour et al (2019) looked at the inverse relationship between amount of caffeine consumed and symptoms of depression; Wang et al (2018) looked at the relationship between the cadmium in the blood and both blood pressure and obesity; Howell et al (2017) looked at the mortality rate of Mexican American adults in relation to their BMI/BMI percentile as well as demographic information.

The focus of the interactive tool as well as this report is to provide summaries of obesity data collected in NHANES. Obesity is a serious problem as documented by the CDC and World Health Organization (WHO) which has led to an increase in many diseases ranging from cardiovascular disease to diabetes worldwide. Hales, Carrol, Fryar, and Ogden (2017) documented the incidence of obesity in the United States based on different demographic data based on a publication within that organization from 2015- 2016. This paper presented this information through bar charts to overall counts by demographic groups as well as line charts to show the incidence rates within the time frame of 2015-2016. They also showed line plots overtime from 1999 to put these results in a general context. Thompson et al (2007) explored the impact of childhood obesity upon later health effects and identified risk factors for long term health. In this study the long term impacts for how the development of cardiovascular disease were examined to see impact upon children as well as effects within

adulthood. The Centers for Disease Control and Prevention looked at incidence of diabetes with regards to many risk factors including obesity (2011). Ford ES (2005) summarized the risk factors for both cardiovascular disease and diabetes and indicated the link between both diseases and obesity.

This interactive tool will provide a summary of the prevalence of obesity using the NHANES data. This will due this for different demographics variables to show the prevalence based on variables like age. Variables were also generated showing the continuous variables like age, height, and weight binned for use in tree or treemap tables. It was felt that from the clinical perspective these were good as they mimicked clinical flow diagrams. Thus this would be good visualization approaches based on how clinicians look at diagnosis strategies based on flow charts.

Data Description

The data included in our combined dataset includes the following:

Summary of Variables				
Variable Name	Units	Type	Labels	Description
Sequence Number		Categorical		Primary key for all csv files for NHANES
Age	Years	Integer		Age at time of the vital sign measurement in years
Age Grouping		Categorical	21-30, 31-40,...	Derived based on ten year intervals
Height	cm	float with 1-decimal		
Height Grouping		Categorical	151-175, 176-200, ...	Derived
Weight	kg	float with 1-decimal		

Weight Grouping		Categorical	71-80, 81-90, ...	Derived
BMI	$\frac{kg}{m^2}$	float with 1-decimal		Derived based on height and weight
Systolic BP	$mmHg$	float with 1-decimal		
Diastolic BP	$mmHg$	float with 1-decimal		
Total Cholesterol	$\frac{mg}{dL}$	integer		
Obesity		Categorical	TRUE, FALSE	TRUE = Obese, FALSE = Not Obese

The data contains the overall key as well as important vital sign variables such as height, weight, BMI, and blood pressure. Some variables have been derived to create categorical groupings. This makes sense as there may be much change year to year for most people under 40 (in the elderly population there would be more impact upon health even from year to year or even month to month). Yet there are significant medical issues which appear from decade to decade as the health problems for a 20 year old are different than a 40 year old. They also allow for these variables to be used in our visualizations such as treemaps. The same is true for the height and weight groupings. The categorical bins allow for the binning of subjects in the space filling approaches to expand to more detailed data on obesity within those groupings.

Data Summary

Data Cleaning

0.2 Description of Interactive Visualization Approach

0.3 *Individual Reports*

Kendall Zettlmeier

The work that I did for the project, started with setting up the team with a GitHub repository where the team could collaborate on the project together and store all of our code and deliverables. I also got the team started with deliverable 0 by creating the text file to be turned in and giving a description of the dataset that the team chose with a brief introduction on what we were going to be doing with the data. I was also able to pull down some data via the Kaggle site to match multiple CSVs into 1 CSV by create an combined excel workbook and matching the survey participants SEQN numbers to each CSV via a VLOOKUP excel function, I then pulled the correct columns in which we wanted to use for the project. As far as the actual implementation of the visualization goes, I was able to get the data setup in such a way to use age, weight, and height as heirarchical datasets in which we could create a treemap of the different data chunks. By splitting the data for age, weight, and height into different groups, it made for visualizing the data into a more managable chunks. I then set the 3 data chunks to be displayed via treemaps via a generic treemap helper that I created, and set the different treemaps to be viewed based upon the radio selection that was chosen. I was also able to get the treemap dispatch function setup to send across the clicked group so that we could hook the click up to open up sub pie charts below the treemap. Finally I was able to collaborate with Kevin on the styling of the web page to align and setup titles for each of the different visualizations.

As far as what I learned from the visualization, is was basically what I expected. The younger the age group the lower the BMI and cholesterol levels were. The younger the age, also had fewer obese survey participants. Since there is a direct correlation to the height and weight to what the BMI was, it wasn't surprising to see that heavier weighted participants had a larger percentage of obesity and high cholesterol. It was on the otherhand interesting to see the differences in height in which groups had higher amounts of obesity. I did not necessarily think that the taller groups would have higher percentages of obesity. I tend to think of most tall people as thin, so it was interesting to see that the tallest group had the highest obesity percentage.

References

- [1] Centers of Disease Control and Prevention. National diabetes factor sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. US Department of Health and Human Services, Centers for Disease Control and Prevention, 1(1), 2568-69.
- [2] Ford ES (2005). Risks for all-cause mortality, cardiovascular disease, and diabetes associated with metabolic syndrome: a summary of the evidence, *Diabetes Care*, 28(7), 1769-1778.
- [3] Hales, CM, Carrol, MD, Fryar, CD, and Ogden, CL (2017). Prevalence of obesity among adults and youth: United States, 2015-2016.
- [4] Howell, CR, Fontaine, K, Ejima, K, Ness, KK, et al (2017). Maximum lifetime body mass index and mortality in Mexican American adults: the National Health and Nutritional Examination Survey III (1988-1994) and NHANES 1999-2010. *Preventing Chronic Disease*, 1545-1151(14).
- [5] Iranpour, S. and Sabour S (2019). Inverse Association between caffeine intake and depression symptoms in US adults: data from National Health and Nutritional Examination Survey (NHANES) 2005-2006. *Psychiatry Research*, 1872-7123(271), 732-739.
- [6] Thompson, DR, Obarzanek, E, Franko, DL, Barton, BA, Morrison, J, Biro, FM, Daniels, SR, Striegel-Moore, RH (2007). Childhood overweight and cardiovascular disease risk factors: the National Heart, Lung, and Blood Institute Growth and Health Study. *The Journal of Pediatrics*, 150(1), 18-25.
- [7] Wang Q, Wei S (2018). Cadmium affects blood pressure and negatively interacts with obesity: Findings in the NHANES 1999-2014. *The Science of the Total Environment*, 1879-1026(643), 270-276.

Appendix Code

If we want to have code or snippets of code in an appendix