Hackathon-IE6600-Group29-Section1

Chaithanya S Gudipati, Pooja Ramesh, Indrayani Yadav

2023-02-12

Introduction

Survey Description

The National Health and Nutrition Examination Survey (NHANES)is a unique combination of interviews and physical examinations designed to assess the health and nutrition status of adults and children in the United States.A key initiative of the National Center for Health Statistics(NCHS) is NHANES, which is a division of the Centers for Disease control and prevention (CDC). This division is compiling national health and vital statistics.

Sample Population and about the Dataset

The sample population consists of non-institutionalized civilians of all ages living in all 50 states and Washington,D.C. Every year the survey examines a nationally representative sample of a approximately 5,000 people, who live in counties across the country. NHANES over-samples people of 60 and older African Americans, Asians and Hispanics to produce reliable statistics.

The 2013- 2014 NHANES dataset includes demographic, examination, dietary, laboratory and questionnaire datasets and provides the information about the demographics socio-economic status, diet and health. Medical professionals with extensive training perform laboratory testing,medical,dental and physiological assessments as a part of the evaluation process.

source:https://www.kaggle.com/datasets/cdc/national-health-and-nutrition-examination-survey?resource=download&select=examination.csv

Additional resources: https://www.cdc.gov/nchs/index.htm

Problem Statement

To determine the prevalence of major disease and disease risk factors in 2013-2014 using the survey's findings. This information will be utilized to evaluate the relationship between nutritional status and the promotion of health considering the national standards for measurements as blood pressure, cholesterol, hypertension along with the lifestyle parameters like smoking, alcohol consumption and work life balance.

We explored the data to understand the underlying factors leading like, demographics, type of environment, nutrition intake and lifestyle leading to the most predominant condition amongst the given population. On further analysis, we will evaluate the following from the data sets :-

- Which is most occurring health condition according to survey conducted by NHANES in 2013-2014? Which age group and race experiences the greatest health issues?

- How this condition contributes to other health conditions from the survey of the population using correlation?
- Evaluating the nutrient consumption and lifestyle that leads to the most prevalent health condition and analyzing the factors that are contributing the most to the disease.
- How the most predominant health condition has varied over the years 1999 to 2018?

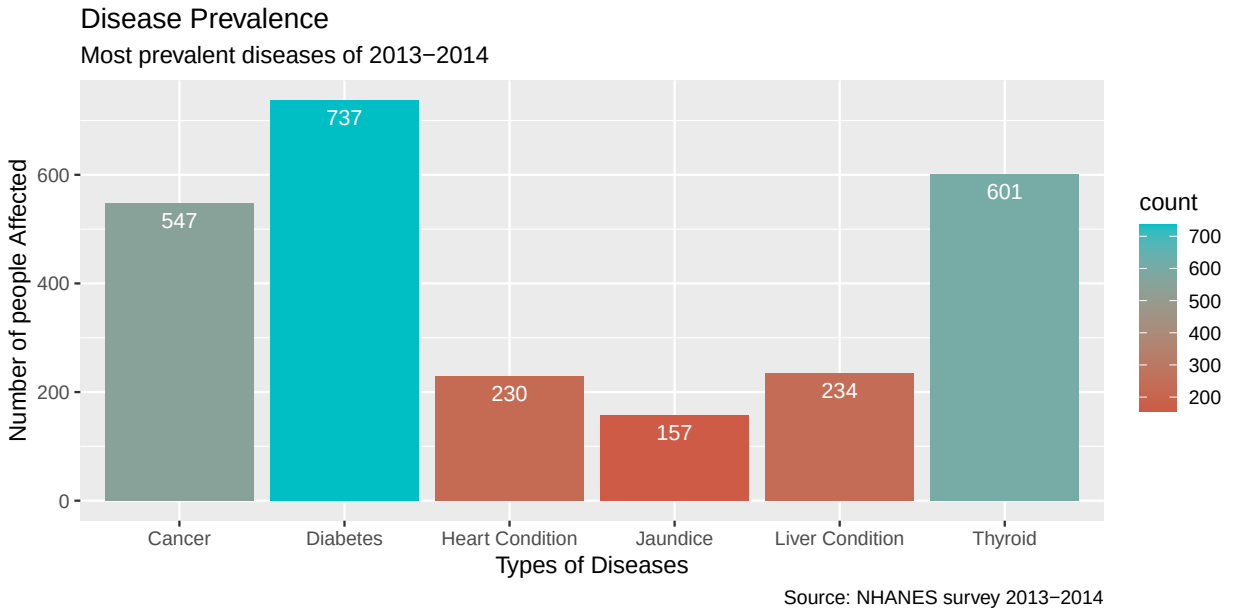
Section 1: Exploring the Demographics and understanding the most occuring health disorder.

Which is most occurring health condition according to survey conducted by NHANES in 2013-2014? Which age group and race experiences the greatest health issues? The respondents in the survey data set consists of 8.8% of Hispanic, 17.1% of Mexican Americans, 10.2% of non-Hispanic Asians, 25.8% of non-Hispanic black, 2.1% of non-Hispanic multiracial and 36.1% of non-Hispanic whites.

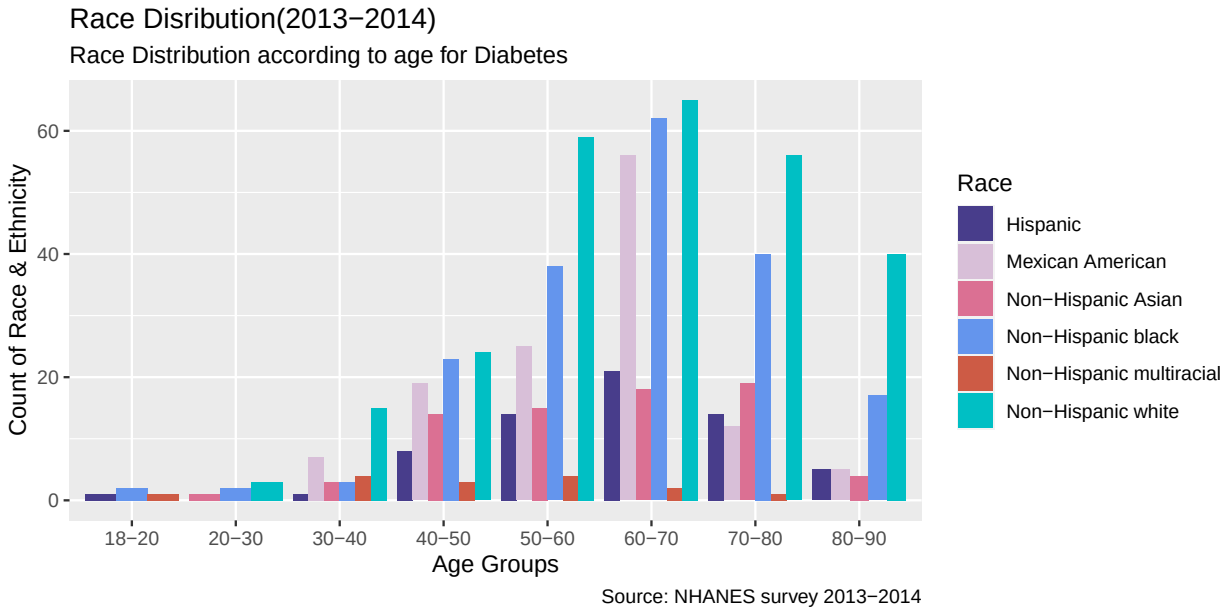
Race	count
Non-Hispanic white	262
Non-Hispanic black	187
Mexican American	124
Non-Hispanic Asian	74
Hispanic	64
Non-Hispanic multiracial	15

In the NHANES data set, there are many prevalent diseases however, we are considering the most common disorders which are cancer, thyroid, diabetes,heart conditions, jaundice and liver conditions affecting the population in the year 2013-2014 according to WHO. From the below figure, we observe that *Diabetes* is the most occurring health disorder, followed by thyroid and cancer according to the census of 2013-2014.

Diabetes is a auto-immune disorder which is chronic(long lasting) and affects how your body turns food into energy. This overtime leads to serious damage to the heart,blood vessels, eyes, kidneys and nerves. According to WHO, about 422 million people worldwide have diabetes and 1.5 million deaths are directly attributed to diabetes each year.

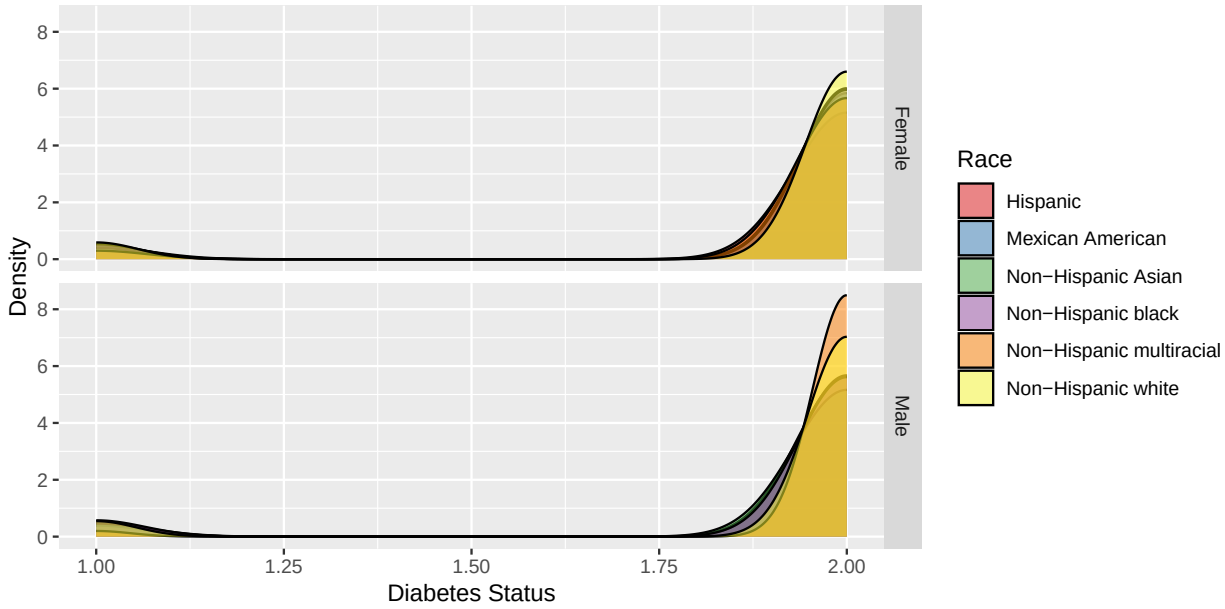


In the above table, we notice that there are 6 different types of Races who were a major part of the survey. Now as we know that diabetes is the most widespread disorder, which race distribution and the age is most affected by diabetes ?

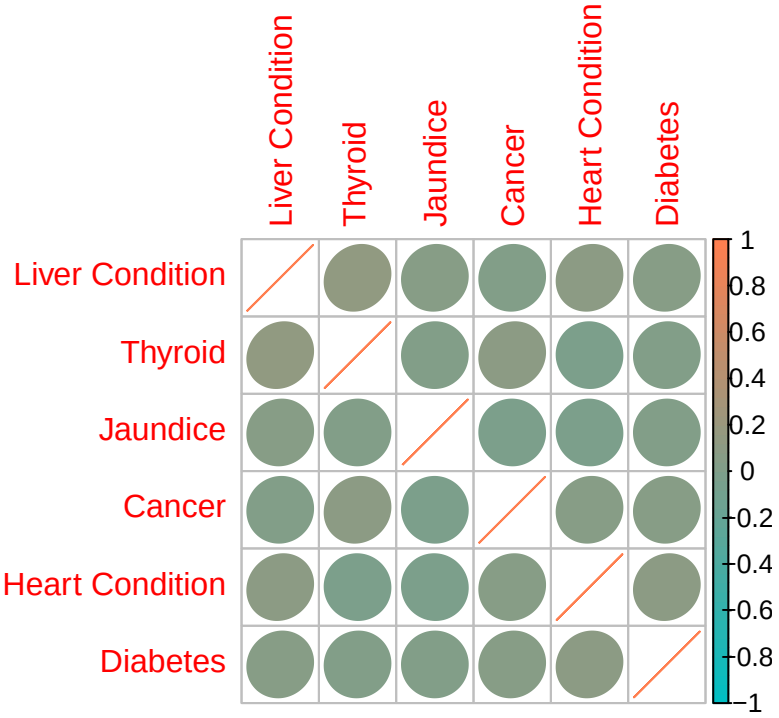


Non-Hispanic whites between 60-70 age range are the most affected with diabetes, meanwhile non-Hispanic multiracial groups are the least affected.Here, we notice that throughout all age groups non-Hispanic whites seems to be the most affected with diabetes, followed by non-Hispanic blacks with very slight less count.

On analysis from the below figure, we can concur that most of the population do not have diabetes, but amongst the ones affected, male population is more.



Correlation plot for Diseases



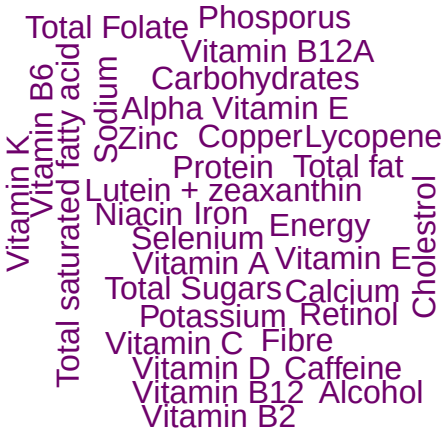
The correlation heat map tells us how presence of diabetes in an individual can effect other disorders from the list.We notice that diabetes does affect other diseases as well highest being *thyroid*.

We can conclude that Diabetes is the most predominant among all the disorders in the survey, followed by thyroid and cancer.On further analysis, we notice that non-Hispanic whites in the ages of 60 to 70 are the most affected population. Among them, male seem to be more affected and prone to Diabetes.

Section 2:

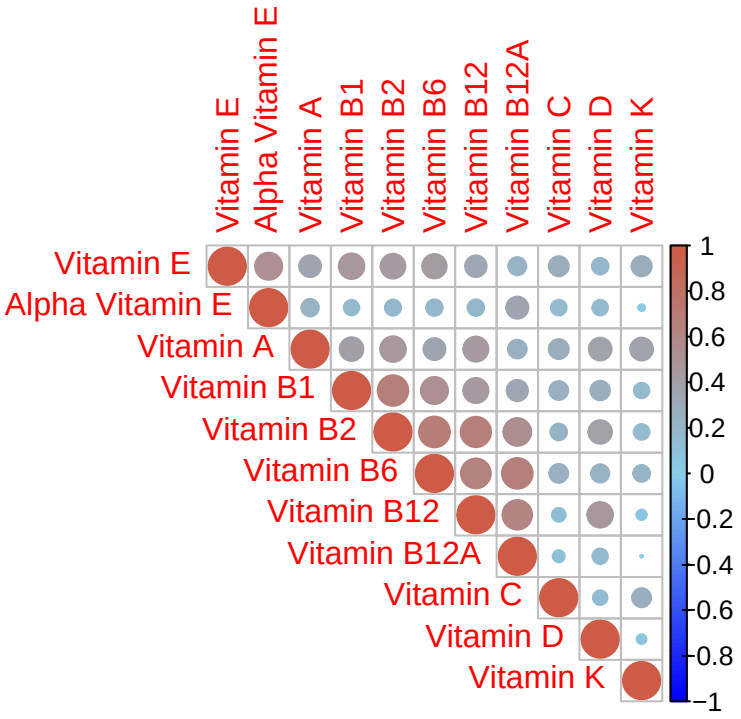
Part 1 : Study of effect of nutrition and lifestyle

This section gives insights into the nutrition and lifestyle that contributes to Diabetes. According to National Center for Biotechnology Information (NCBI), we have considered the below list of nutrients (minerals & vitamins), whose consumption greatly leads to diabetes.

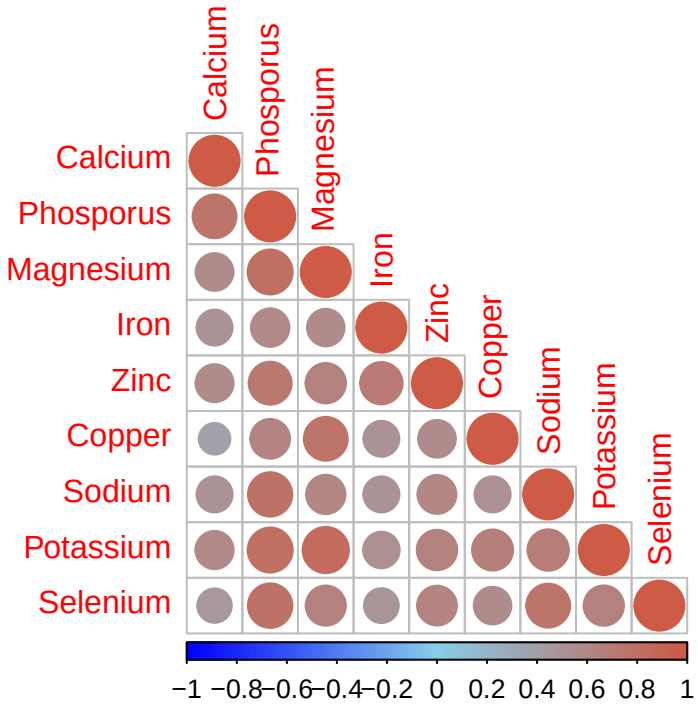


Evaluating the nutrient consumption and lifestyle that leads to the most prevalent health condition and analyzing the factors that are contributing the most to the disease.

Correlation plot for Vitamins



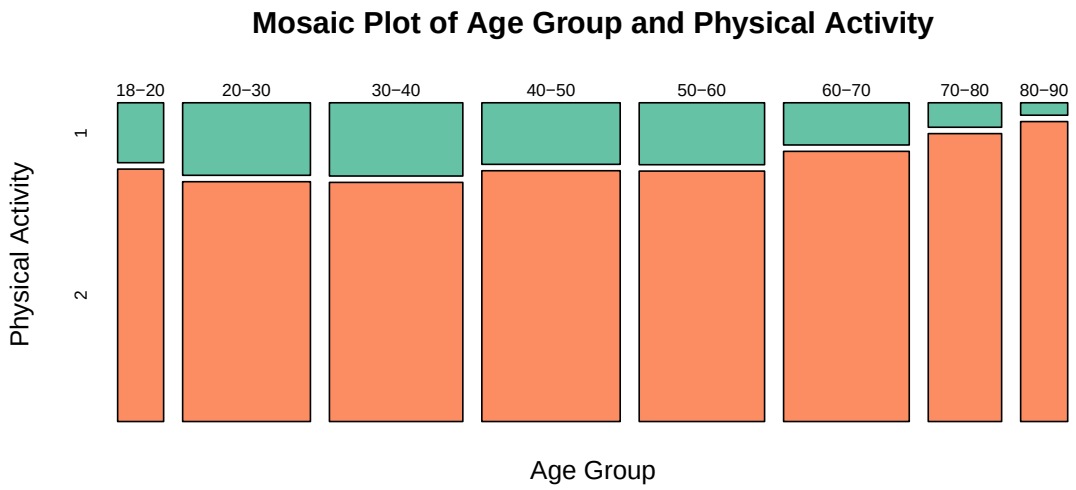
Correlation plot for Minerals



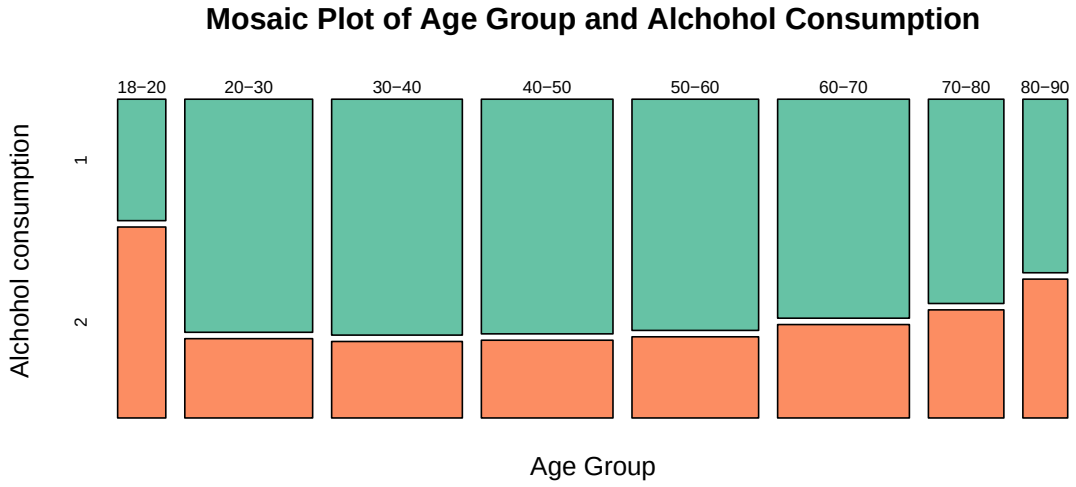
According to WHO, deficiency of magnesium is the main cause for diabetes. From the heat maps, it is clearly evident that the correlation between Magnesium and phosphorous have the highest relation of 0.85, leading us to conclude that they do have the highest affect on Diabetes. Furthermore, we can conclude that Vitamins B6,B2 and B12 lead to this disorder.

Lifestyle Factors - Alchohol consumption and Physical Activity

There was recent article on MedicalNewsToday which highlighted the main lifestyle changes which need to be incorporated in order to reduce diabetes.Great emphasis was laid on Physical activity and Alcohol consumption.



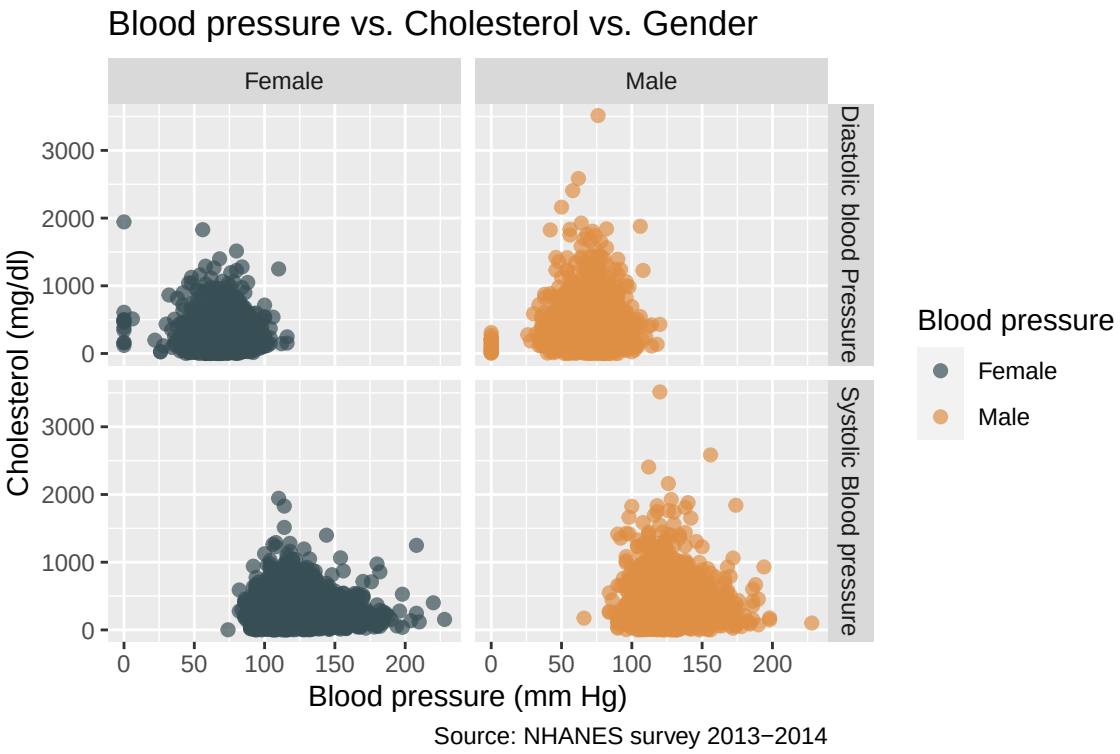
From the mosaic plot, it is evident that people who do rigorous physical activities are least affected by diabetes.

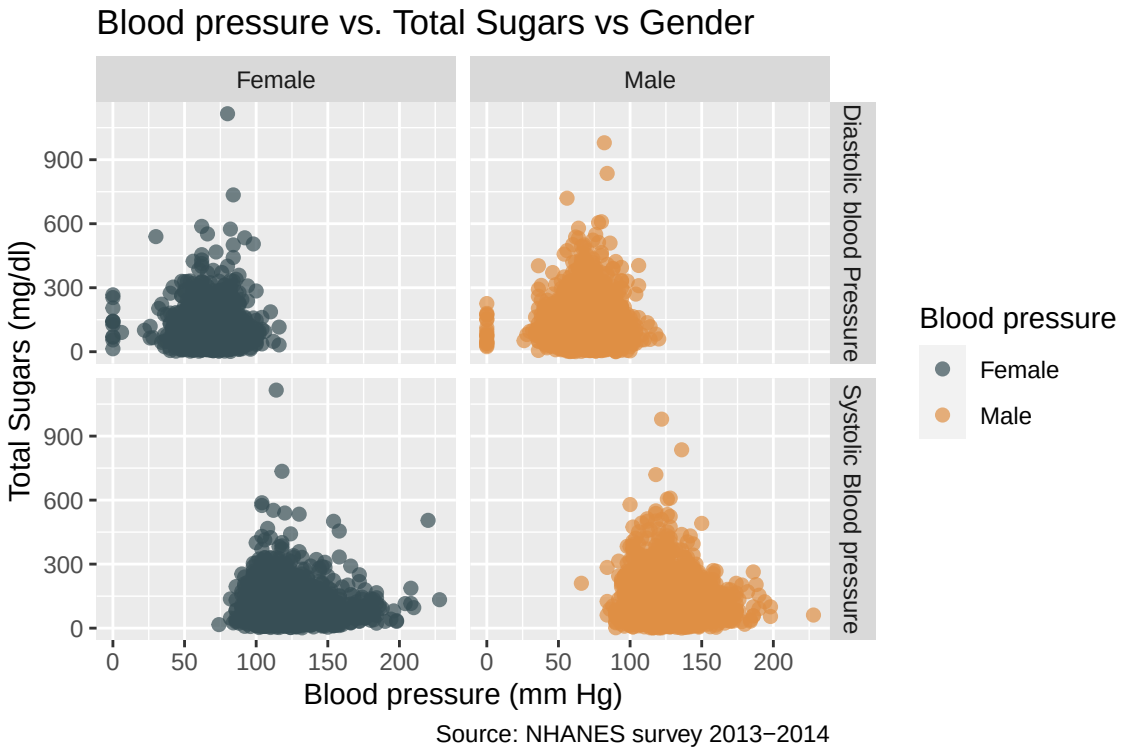


Here, the graph concludes alcohol consumption does affect diabetes greatly. People who consume alcohol are more prone to being diabetic and from this plot, we can also conclude that alcohol intake is more among the age group 20-70.

Part 2 : Effect of blood sugar levels, chorestrol and hypertension

Poor blood sugar levels can worsen your cholesterol levels. A diabetic patient usually has low levels of good cholesterol and high levels of bad cholesterol.

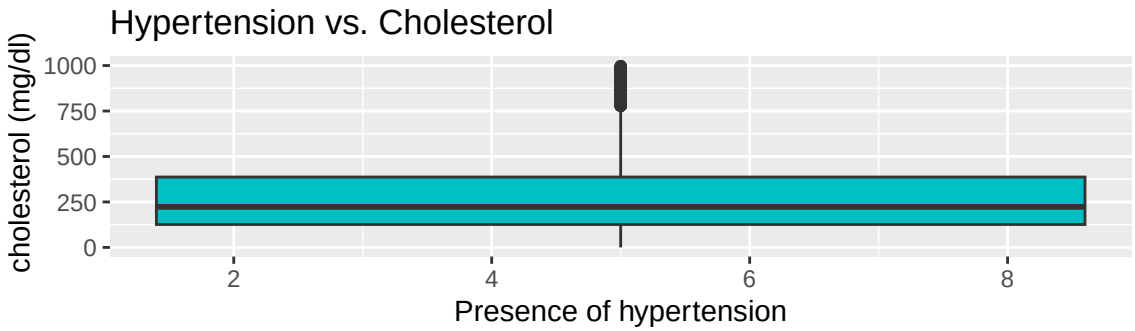




Majority of the female diabetic patients have sugar levels between 0 and 300 mg/dl and a diastolic Blood pressure of range 50-100 mm Hg whereas for a male patient, the blood pressure remains similar, howsoever the sugar levels vary between 0-400 md/dl. Systolic blood pressure is similar for both ranging from 75-

175 mm Hg.We can observe similar trends for cholesterol as well where male has slightly higher values in comparison to the female.

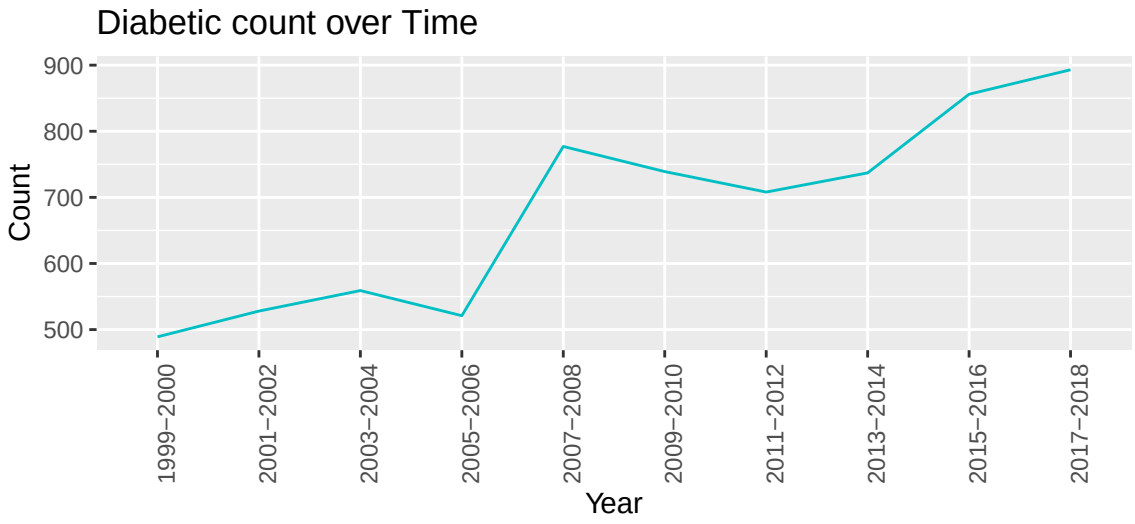
Even though we noticed that work sectors does not affect the onset of diabetes, however, stress or hypertension is major factor in this. Hypertension leads to high levels of cholesterol which is in turn is a major cause for the onset of diabetes in an individual.From the below plot, 75 percentile of individuals who have diabetes have a cholesterol level ranging from 125 - 375 mg/dl with a median value of 225 mg/dl.



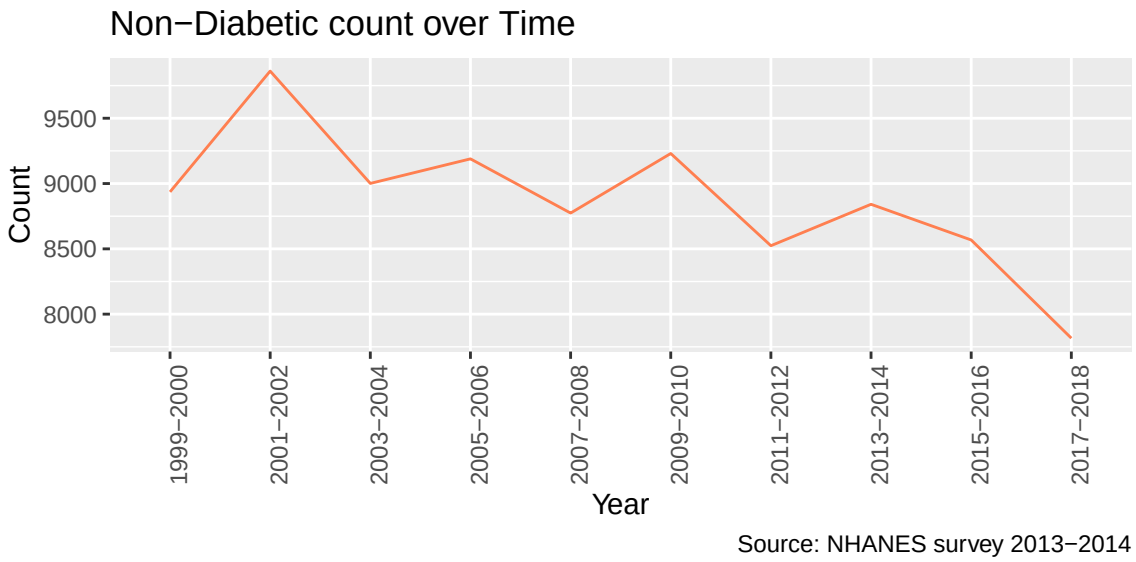
Source: NHANES survey 2013−2014

Section 3

How the most predominant health condition has varied over the years 1999 to 2018?



Source: NHANES survey 2013−2014



The above plots are based on diabetes data from 1999-2018.Diabetes trends have varied greatly over the years and we can see a clear increase in count over the years. We have further analysed that there is drastic increase in count between the years 2006 and 2007 followed by slight variation over the years. Furthermore, we have observed the non diabetic trend as well.

Conclusion

Over this research, we have learnt Diabetes disorder is the most occurring condition which have drastically changed over the years from 1999-2018.Our data-set was a vast data-set with over 2500 columns and we encoded and chose the columns specific to our analysis for diabetes.Through the analysis, we got insights into how demographic, nutrition,lifestyle and workplace factors affect various disorders, here in specific Diabetes. Starting with the demographics, we noticed that the Non-Hispanic whites (male) were the most affected with diabetes 2013-2014.

We further delved deep into the various nutrition and lifestyle factors that affect diabetes. On research from WHO and prominent medical papers,we were able to project and understand the important minerals and vitamins that have impact on diabetes.We analysed that magnesium and vitamin B12,B6 and B2* are the contributing nutrients.Moving to the lifestyle factors, physical activity and alcohol assumption had greater effect on diabetes.Even though the work place sectors play no role for the onset of diabetes, the stress or hypertension levels seem to have a great effect as they do greatly vary our blood sugar and cholesterol levels.From the time series plots,we observe that both the number of cases and the prevalence of diabetes have been steadily increasing over the past few decades.This survey's data will be utilized in epidemiological studies and health sciences research, which aid in the creation of solid public health policies, the direction and planning of health programs and services, and the advancement of national health literacy.

References

-https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/diabetes-long-term-effects

-https://www.who.int/news/item/14-11-2022-insulin-and-health-technology-manufacturers-make-commitments-in-support-of-who-asks -https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7353202/#:~:text=Low%20magnesium%20intake

-https://www.cdc.gov/nchs/nhanes/index.htm