Kidney Stone Disease

Group 2

2024-08-29

Contents

Background and Data

Dataset

This project is based on the National Health and Nutrition Examination Survey (NHANES) from the National Center for Health Statistics, of the Centers for Disease Control and Prevention. Data from the most recent cycle is used, NHANES 2017 - March 2020.

NHANES is an ongoing program of surveys in the United States that assesses the health and nutritional status of adults and children. The surveys collect health-related data ranging over a number of topics, which are organised broadly into Demographics, Dietary, Examination, Laboratory, and Questionnaire. *explain why dataset is of interest + what questions it could be used to answer + what question we chose to answer*

Data Structure and Types

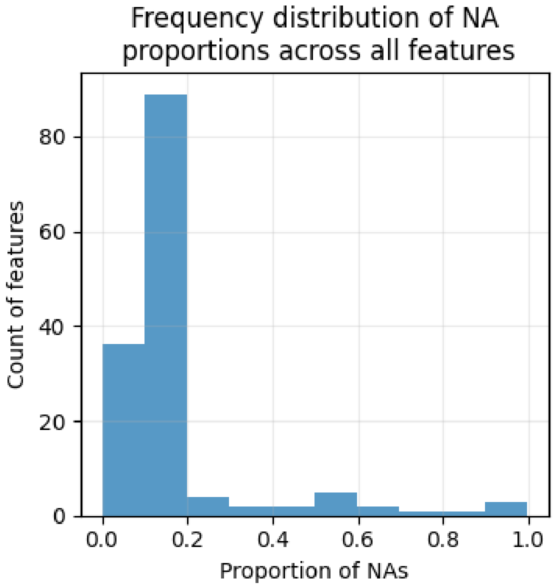
Data from each NHANES cycle is released as many tables, each containing a collection of similar features. For the specific focus on kidney stone disease, only a subset of tables was used, and from these tables, only a subset of key features. The integrated dataset used in this project is composed of 57438 instances/rows, and 146 columns. The column `SEQN` contains a unique identifier for each instance, and the column `KIQ026` contains the target variable. Thus, there are 144 informative features.

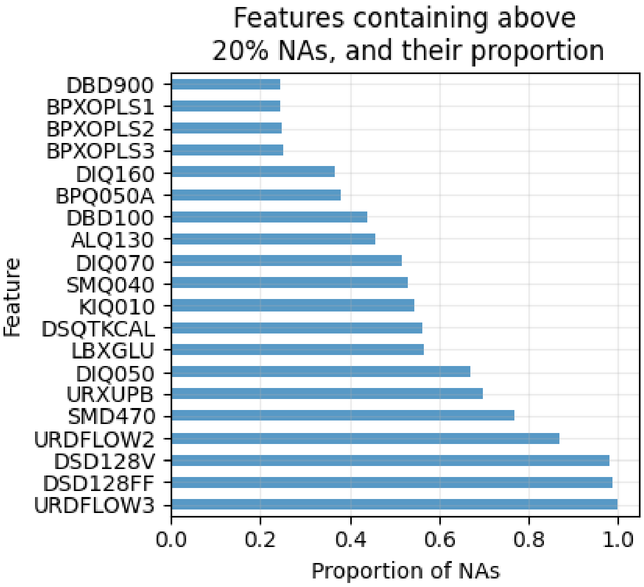
The key features are broadly described in the following:

- Demographic: gender, age, race, education, marital status, and income. Men and older individuals are more likely to have had kidney stones *citation*, and there is evidence that kidney stone prevalence and severity is associated with various socioeconomic factors *citation*.
- Dietary: vitamin, water, nutrient, and dietary supplement intake. Kidney stone incidence increases with certain dietary habits, such as low calcium, low potassium, and low fluid diets *citation*. Everyday foods in the NHANES dietary interviews are deconstructed and aggregated into their nutritional components, thus there is highly specific (and largely correlated) dietary and nutrient data that constitutes a significant portion of the total features explored.
- Examination: body mass index (BMI), blood pressure, and pulse readings. Indicators of general health are useful predictive features for kidney stone risk *citation*.
- Laboratory: aspects of biochemistry profile, and urine-associated tests. Detection of kidney diseases or urinary tract abnormalities (that can lead to kidney stones) are often tested by assessing levels of components such as glucose *citation*, lead *citation*, and the albumin creatinine ratio *citation* in urine.
- Questionnaire: past medical history (conditions and medicines), dietary and alcohol habits, urinary tract function, physical activity, smoking, and sleep habits. Again, general health, behaviours, and lifestyle have a large influence on kidney stone disease. Factors such as lack of physical activity and smoking can indirectly damage the urinary tract and promote stone formation *citation*.

Feature type ranges from numerical continuous to catgeorical ordinal, nominal, and binary. Dietary, examination, and laboratory data are mainly numerical, while demographic and questionnaire data are mainly categorical. To avoid difficult or complicated natural languange processing or text mining, free-text data was not selected.

Data Completeness





Data Integration (if applicable)

[If you used multiple datasets, explain how you integrated them.]

Ethics, Privacy and Security

Ethical Considerations

[Discuss any ethical considerations relevant to your project.]

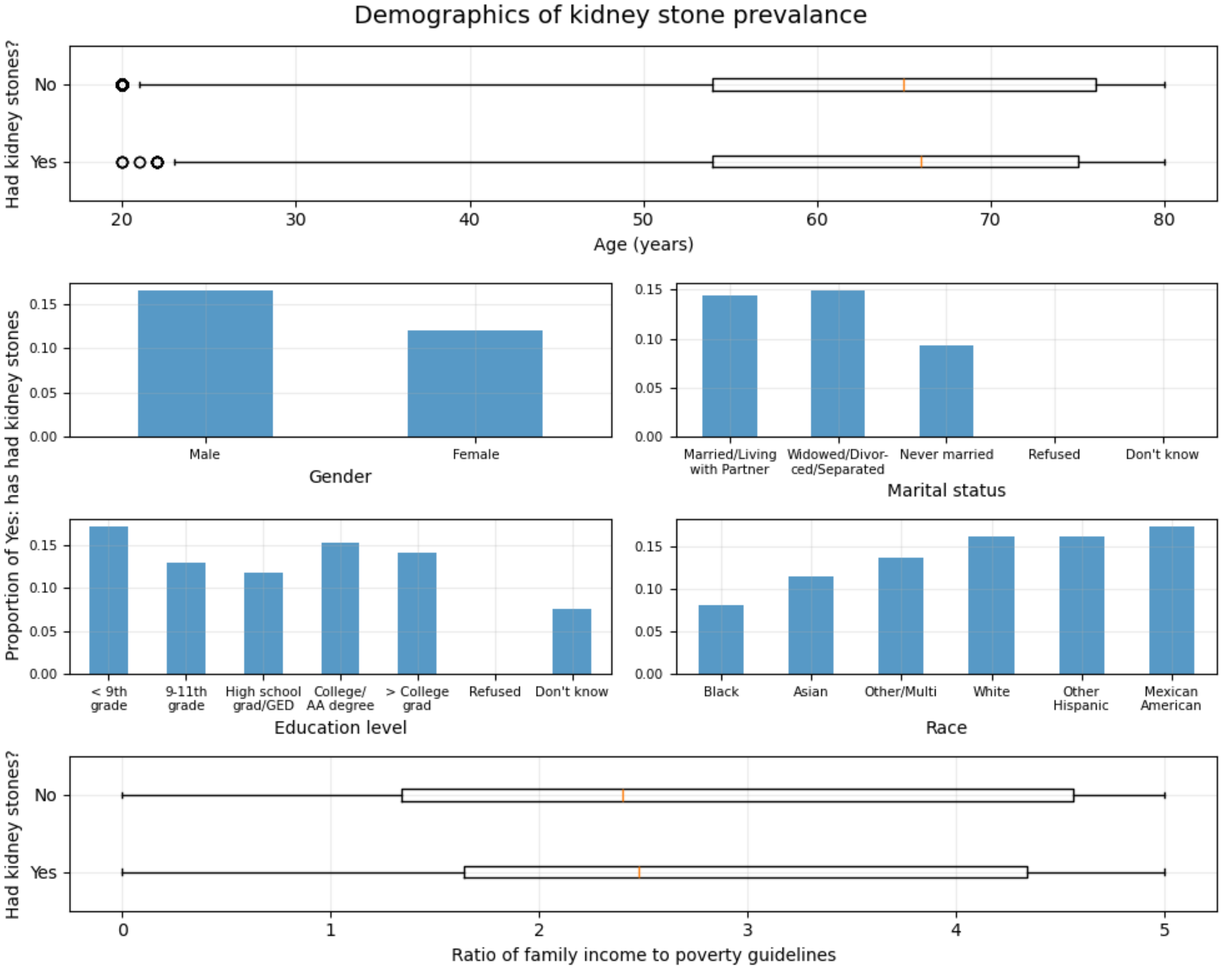
Privacy Concerns

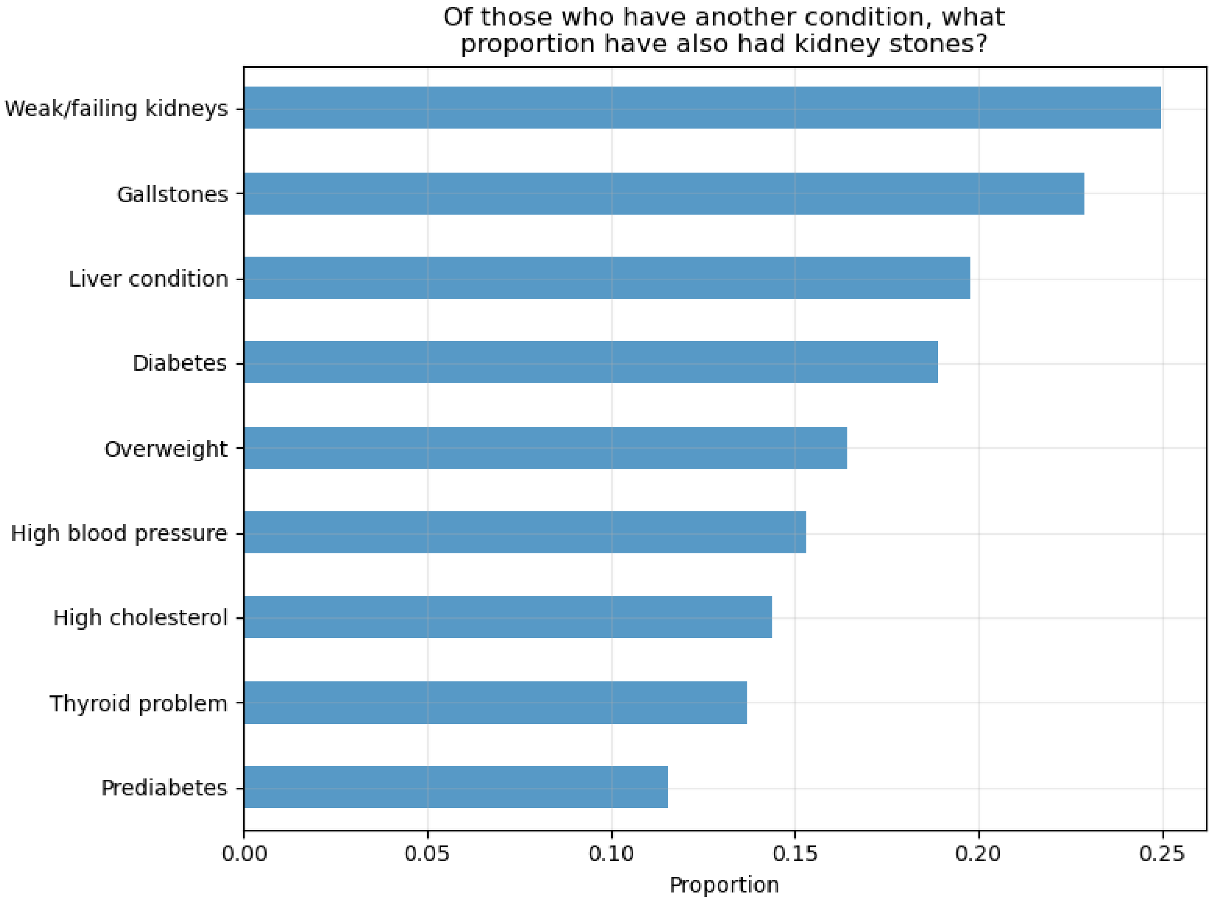
[Address any privacy concerns related to your project.]

Security Measures

[Explain potential steps to keep your project data and results secure. Distinguish between actual steps taken and hypothetical measures.]

Exploratory Data Analysis

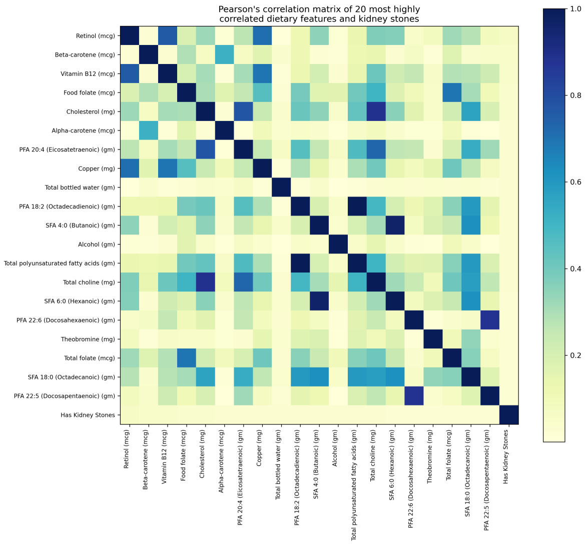


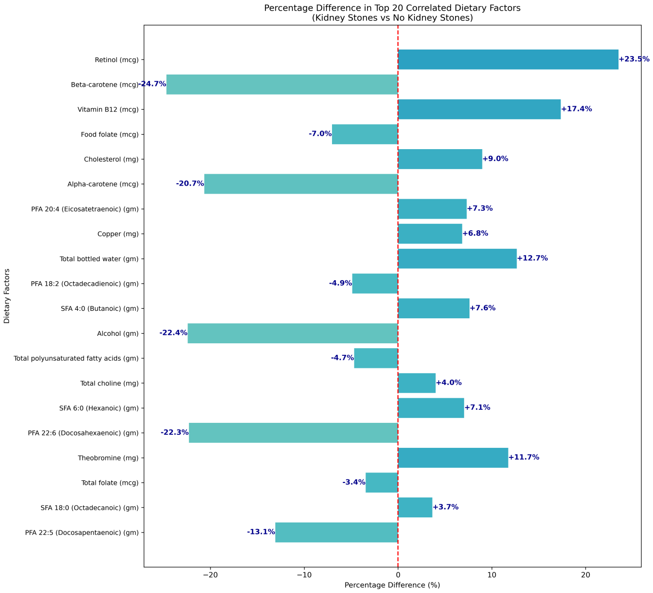


[Explain the variables, comment on the main features, and interpret the results.]

Correlation and Percentage Difference Analysis

To explore the relationship between dietary factors and kidney stones, we conducted a correlation analysis and calculated the percentage difference of these factors between individuals with and without kidney stones. The following graph shows the results of this analysis:





[Here, provide an interpretation of the graph. Explain what the correlation matrix shows, what the percentage differences indicate, and discuss any notable patterns or findings from this analysis. You may want to highlight the most strongly correlated dietary factors with kidney stones and the largest percentage differences between those with and without kidney stones.]

Analysis 2: [Title]

[Explain the variables, comment on the main features, and interpret the results.]

Analysis 3: [Title]

[Explain the variables, comment on the main features, and interpret the results.]

Individual Contributions

[State the contributions of each group member to data preparation, analysis, and report writing.]

References

[List your references here using your preferred citation style.]