Kidney Stone Disease

Group 2

2024-09-03

Contents

1. Introduction

[Brief introduction to the project and its objectives]

2. Background and Data

2.1 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.

[Add: Explanation of why this dataset is of interest, what questions it could be used to answer, and what specific question this project aims to address]

2.2 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.

[Add: More detailed explanation of feature types and their relevance to kidney stone disease]

2.3 Data Completeness

27 features have no missing values (not including the unique identifier and target variable columns). Features that do have missing data can be summarised as follows:

- 100 features have under 25% missing data;
- 4 features have 25 - 50% missing data;
- 7 features have 50 - 75% missing data;
- 5 features have 75% - 100% missing data.

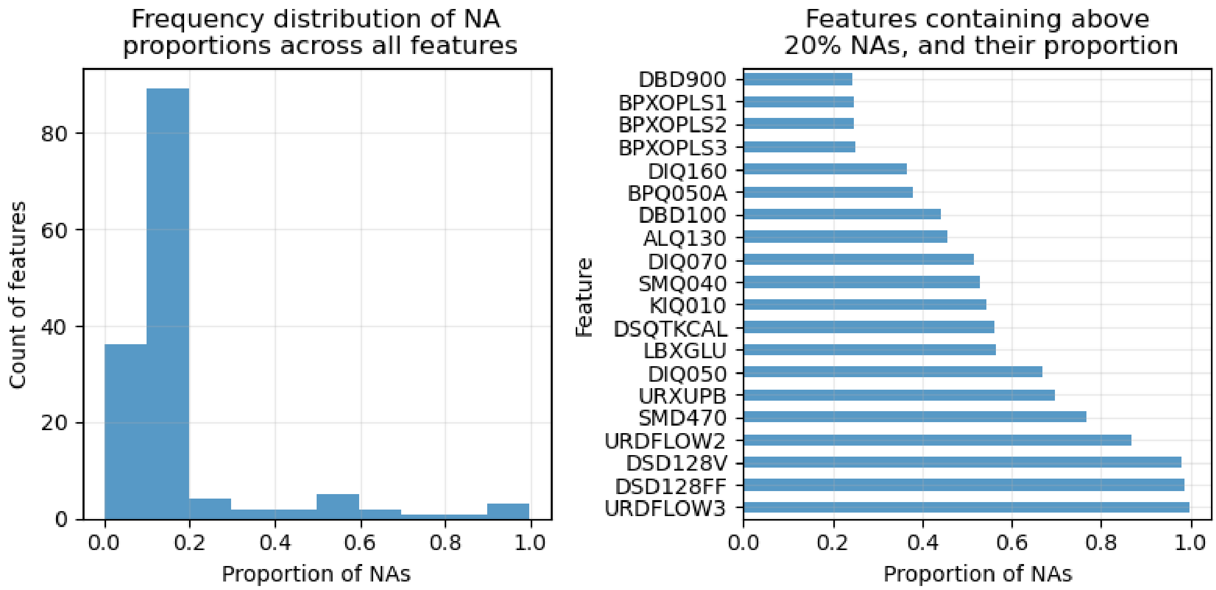

Figure 1: Distribution of missing data across features

3. Ethics, Privacy, and Security

3.1 Ethical Considerations

[Discuss ethical considerations relevant to your project, such as potential biases in the data or implications of findings]

3.2 Privacy Concerns

[Address privacy concerns related to your project, such as handling of personal health information]

3.3 Security Measures

[Explain actual and potential steps to keep your project data and results secure]

4. Methodology

[Describe the methods used for data cleaning, preprocessing, and analysis]

5. Exploratory Data Analysis

5.1 Demographic Analysis

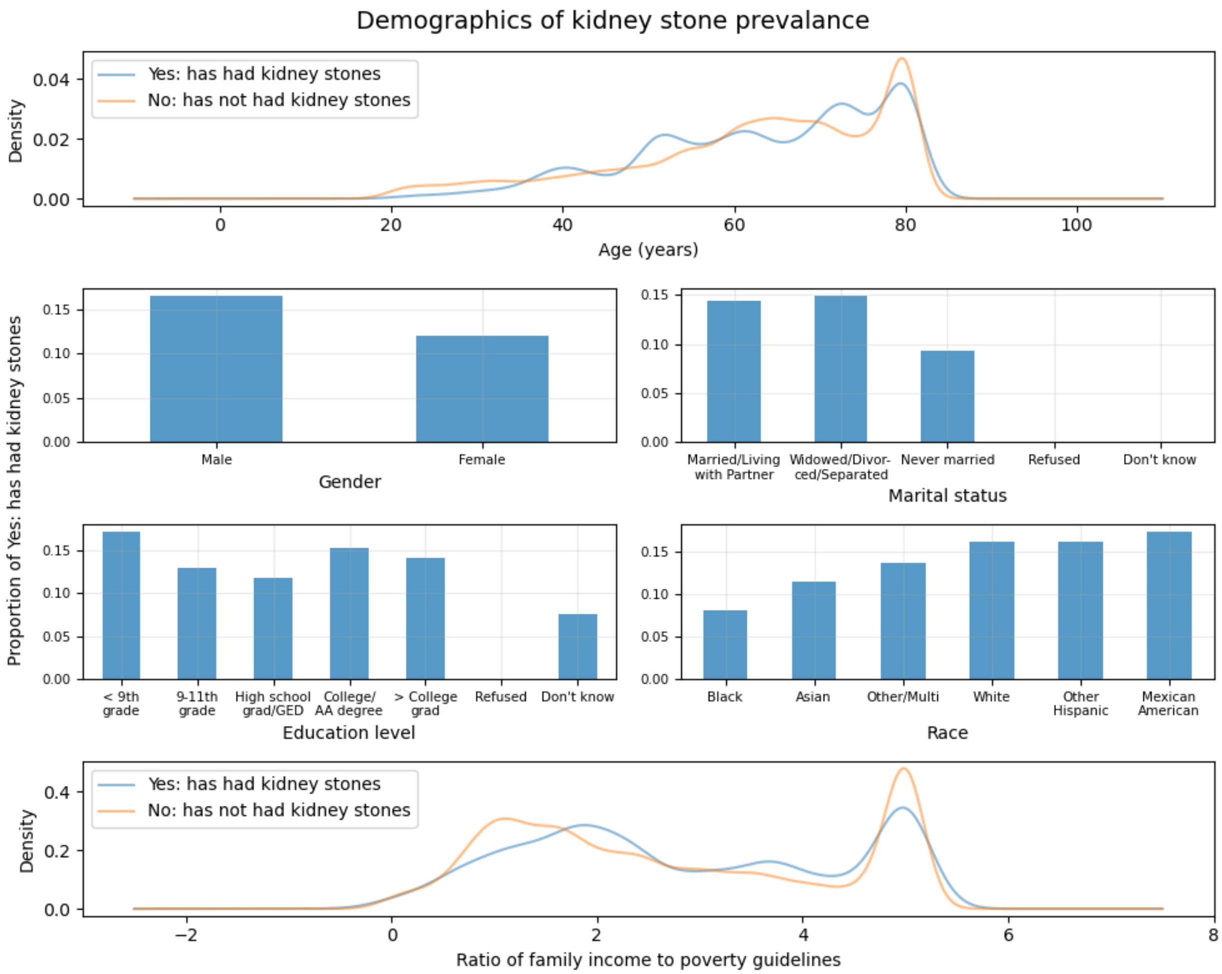

Figure 2: Demographic Kidney Stone Prevalence

[Detailed description and interpretation of the demographic kidney stone prevalence figure]

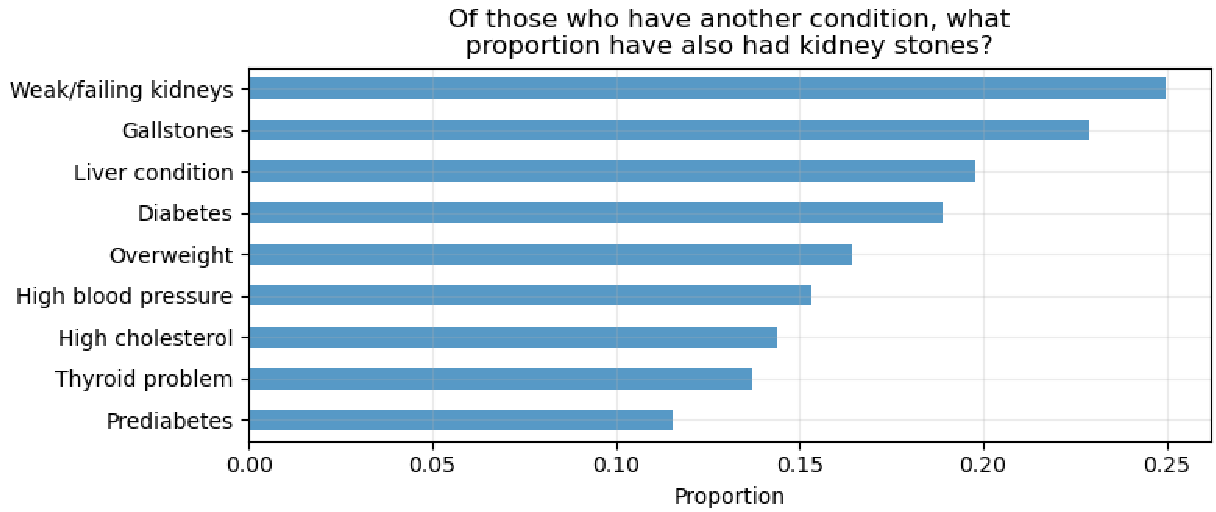

Figure 3: Condition Kidney Stone Proportion

[Detailed description and interpretation of the condition kidney stone proportion figure]

5.2 Laboratory Analysis

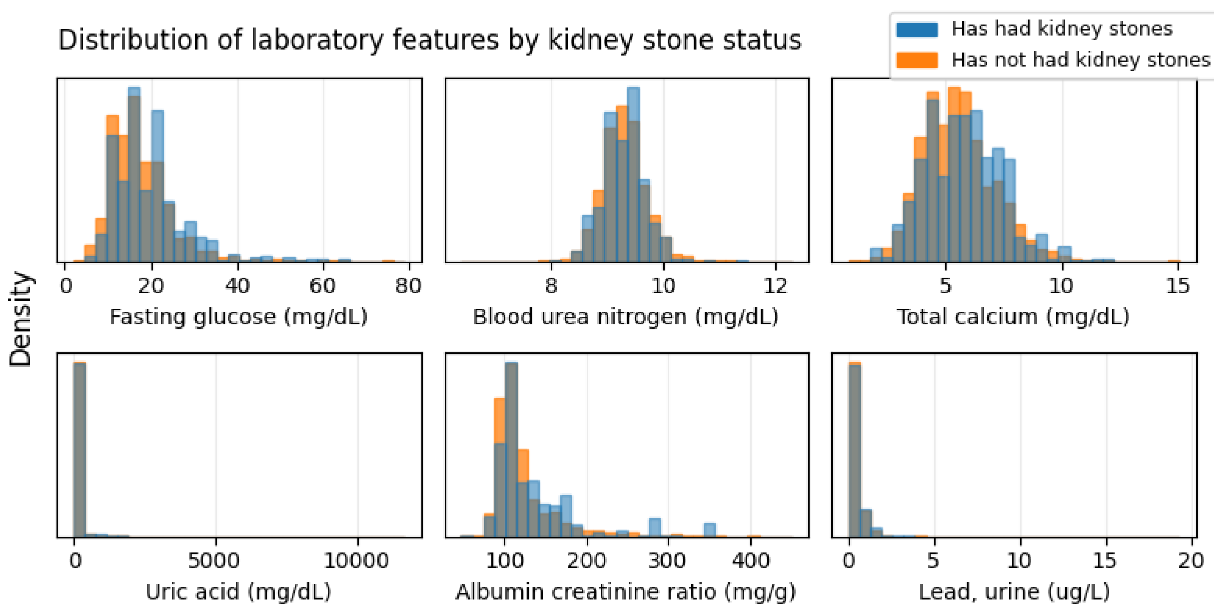

Figure 4: Laboratory Distribution

[Detailed description and interpretation of the laboratory distribution figure]

5.3 Dietary Analysis

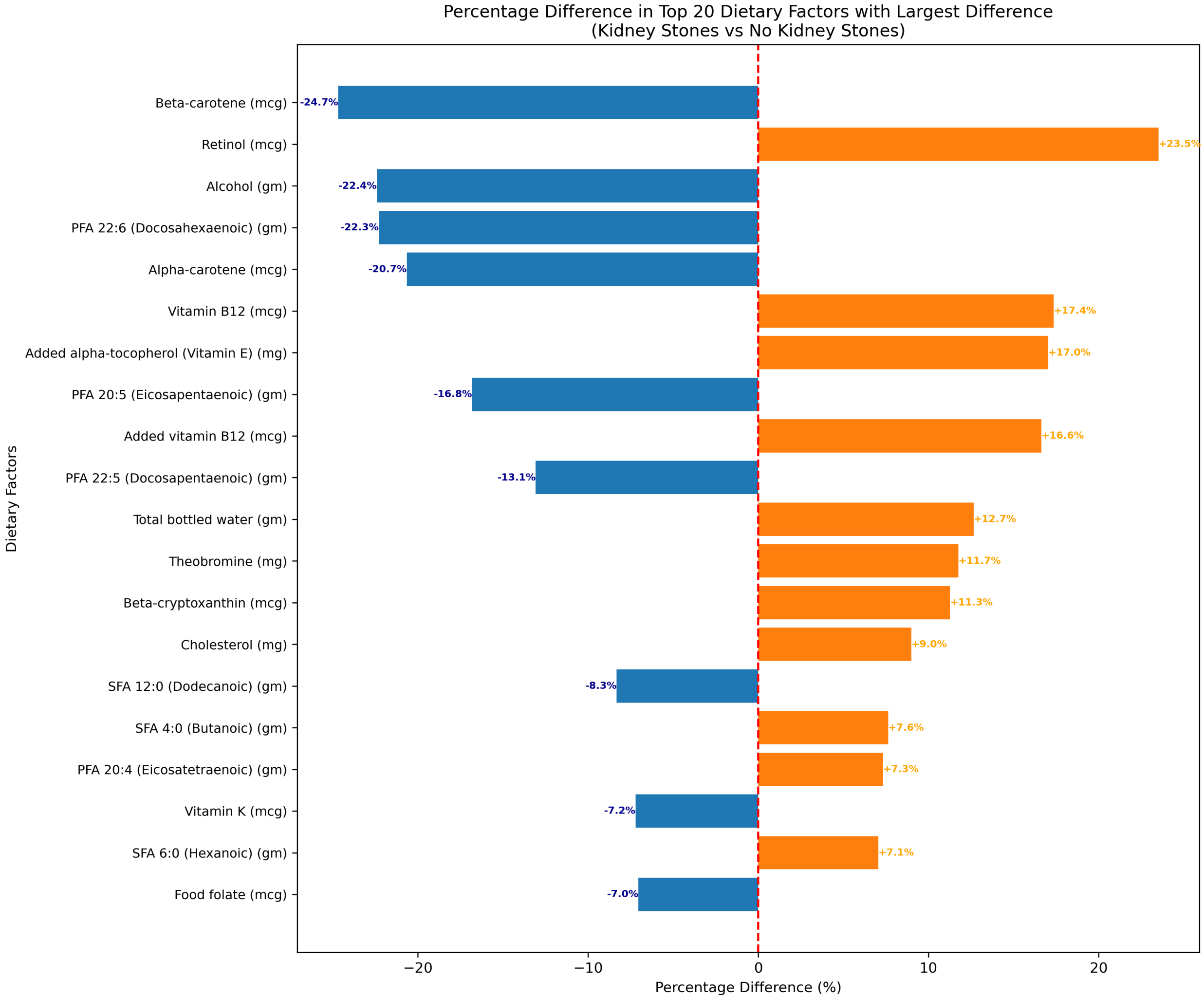

Figure 5: Dietary Differences

Retinol (a form of vitamin A) shows the largest positive difference, with individuals who have had kidney stones consuming approximately 23.5% more than those who have not. This suggests a potential association between higher retinol intake and a history of kidney stones.

Conversely, beta-carotene (another form of vitamin A) displays the most substantial negative difference, with those who have had kidney stones consuming about 24.7% less. This unexpected finding warrants further investigation into the potential protective effects of beta-carotene or differences in vitamin A metabolism among those with a history of kidney stones.

Among the top factors, we see a trend in vitamins and antioxidants, particularly forms of vitamin A, vitamin B12, and vitamin E (alpha-tocopherol). This pattern suggests that the balance and forms of certain vitamins may be different in individuals with a history of kidney stones. Interestingly, alcohol consumption shows a large negative difference (-22.4%), indicating that individuals who have had kidney stones tend to consume significantly less alcohol. This could suggest that those with a history of kidney stones may have modified their diet to reduce alcohol intake. The substantial differences observed in polyunsaturated fatty acids (PFAs), particularly docosahexaenoic acid (DHA, -22.3%) and eicosapentaenoic acid (EPA, -16.8%), indicate that these dietary components might be particularly important in distinguishing between individuals with a history of kidney stones and those without.

Total bottled water consumption is 12.7% higher in those with a history of kidney stones, which could reflect increased fluid intake as a preventive measure.

Cholesterol intake is 9% higher in individuals who have had kidney stones, which might be relevant to kidney stone formation. ## 5.4 Physical Activity Analysis

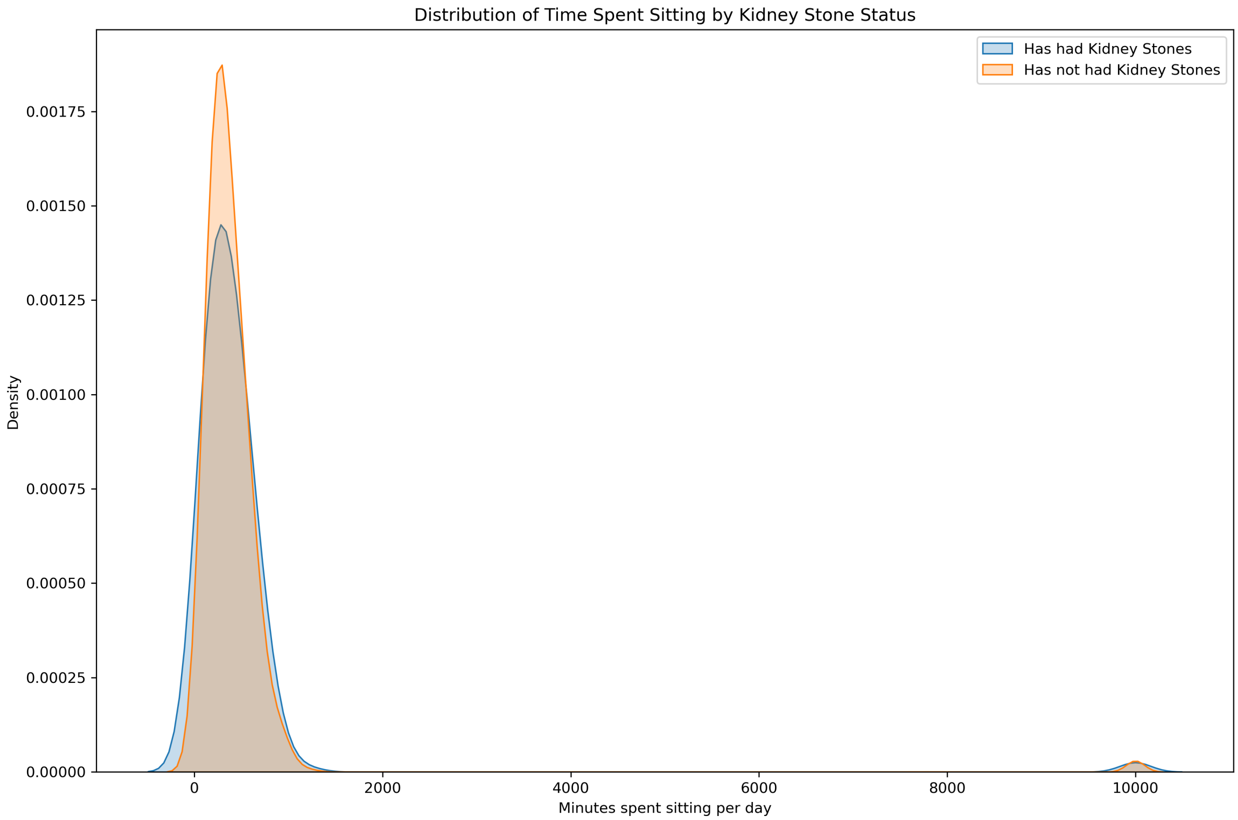

Figure 6: Time Spent Sitting

Both groups show a similar overall pattern, with the majority of individuals spending between 0 and 500 minutes (approximately 0-8.33 hours) sitting per day. However, there are notable differences: those without kidney stones (orange line) have a slightly higher peak density at lower sitting times, suggesting they are more likely to spend less time sitting overall. In contrast, the distribution for those with kidney stones (blue line) is slightly flatter and shifted slightly to the right, indicating a tendency towards longer sitting durations. Interestingly, both groups show a small secondary peak around 9000-10000 minutes (150-167 hours) per day, which likely represents outliers or potential data collection errors, as these values exceed the number of minutes in a day.

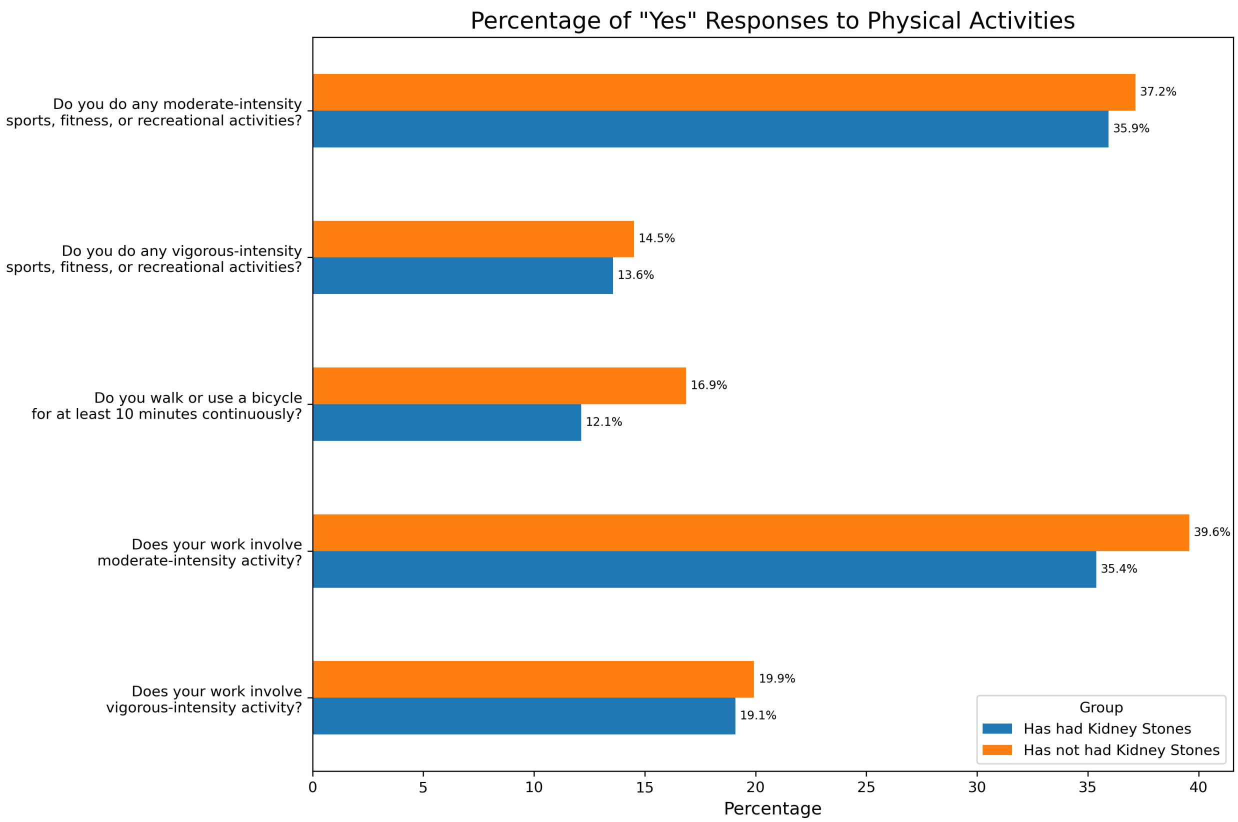

Figure 7: Physical Activities

Across all categories of physical activity, individuals without a history of kidney stones consistently reported slightly higher levels of engagement. The largest differences were observed in Walking or cycling for at least 10 minutes (4.8 percentage point difference). The smallest difference was in work-related vigorous-intensity activity (0.8 percentage point).

These findings suggest a possible association between higher physical activity levels and lower likelihood of kidney stone formation. However, the differences are relatively small, and causation cannot be inferred from this data alone. # 6. Discussion

[Summarize key findings and their implications] [Discuss limitations of the study] [Suggest areas for future research]

7. Conclusion

[Provide a concise summary of the main findings and their significance]

8. Individual Contributions

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

9. References

[List references using a consistent citation style]