

# Kidney Stone Disease

Group 2

2024-09-08

## Contents

<b>1. Introduction</b>	<b>2</b>
<b>2. Background and Data</b>	<b>2</b>
2.1 Dataset . . . . .	2
2.2 Data Structure and Types . . . . .	2
2.3 Data Completeness . . . . .	2
<b>3. Ethics, Privacy, and Security</b>	<b>3</b>
3.1 Ethical Considerations . . . . .	3
3.2 Privacy Concerns . . . . .	3
3.3 Security Measures . . . . .	3
<b>4. Methodology</b>	<b>3</b>
<b>5. Exploratory Data Analysis</b>	<b>4</b>
5.1 Demographic Analysis . . . . .	4
5.2 Laboratory Analysis . . . . .	5
5.3 Dietary Analysis . . . . .	6
5.4 Physical Activity Analysis . . . . .	7
<b>6. Discussion</b>	<b>10</b>
<b>7. Conclusion</b>	<b>10</b>
<b>8. Individual Contributions</b>	<b>10</b>
<b>9. References</b>	<b>10</b>

# 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 9208 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:

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

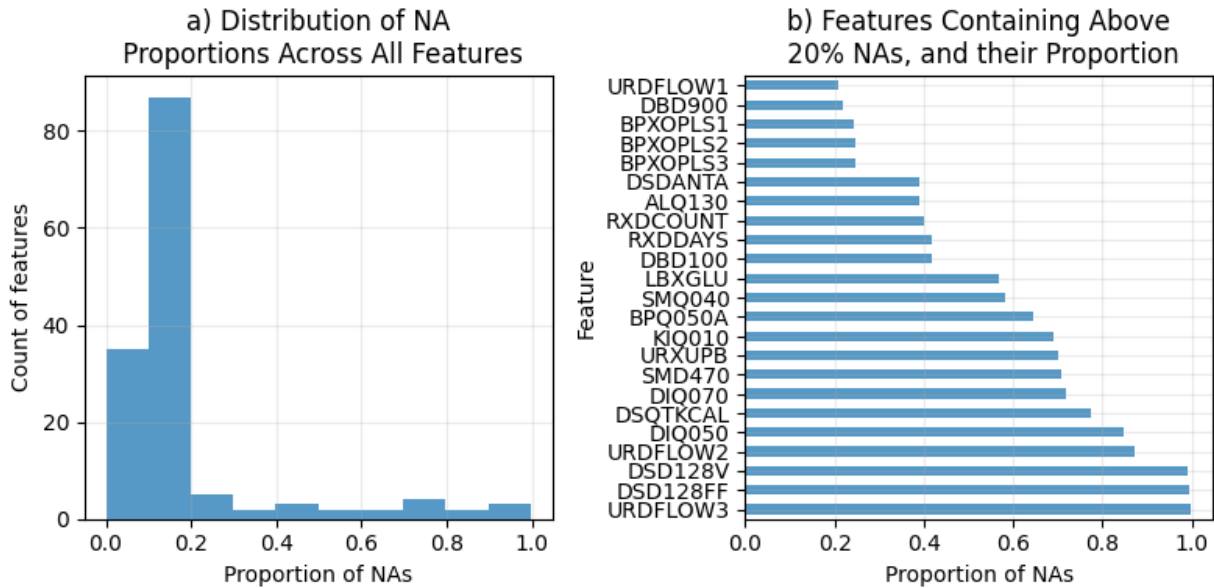


Figure 1: Distribution of missing data across features

### 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

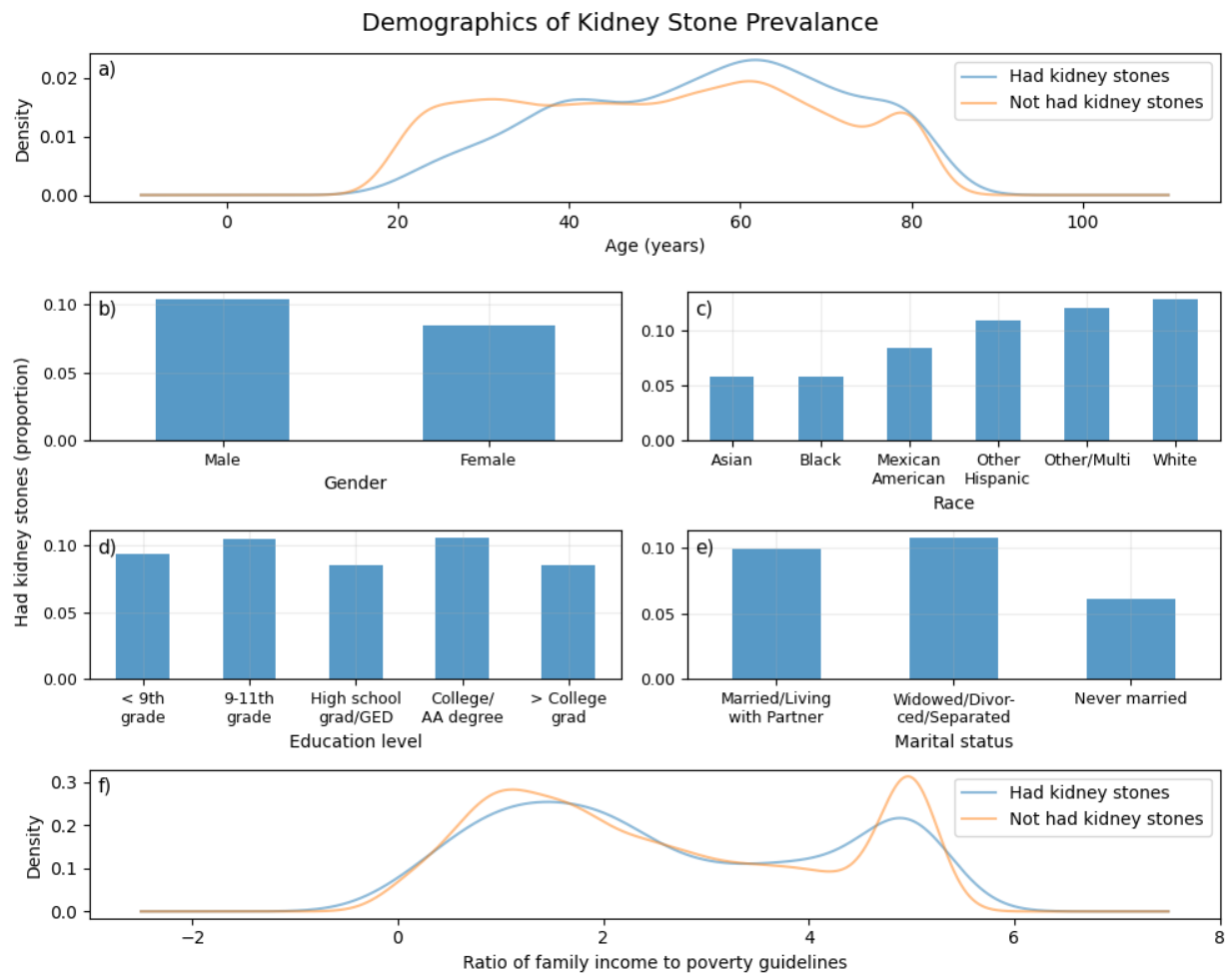


Figure 2: Demographic Kidney Stone Prevalence

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

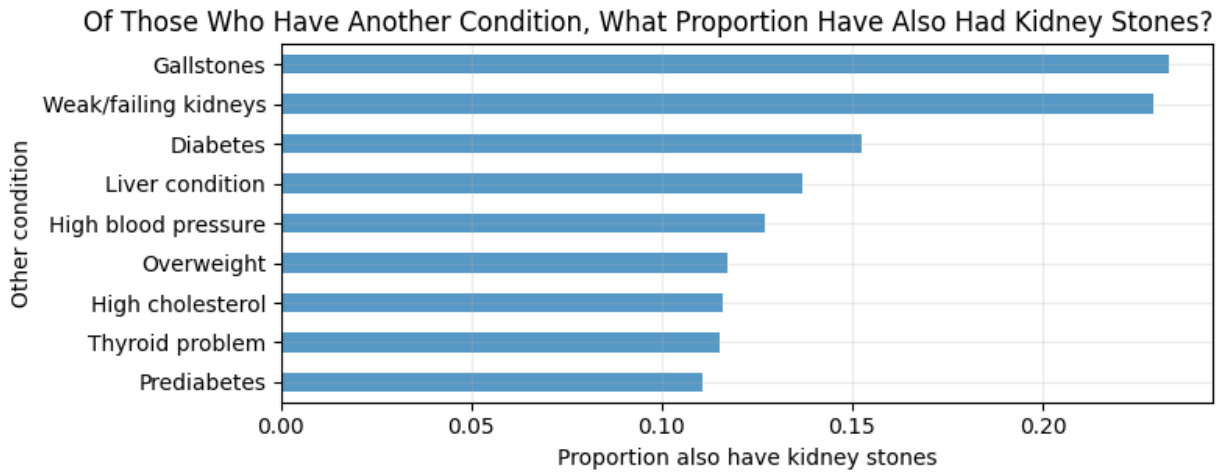


Figure 3: Condition Kidney Stone Proportion

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

## 5.2 Laboratory Analysis

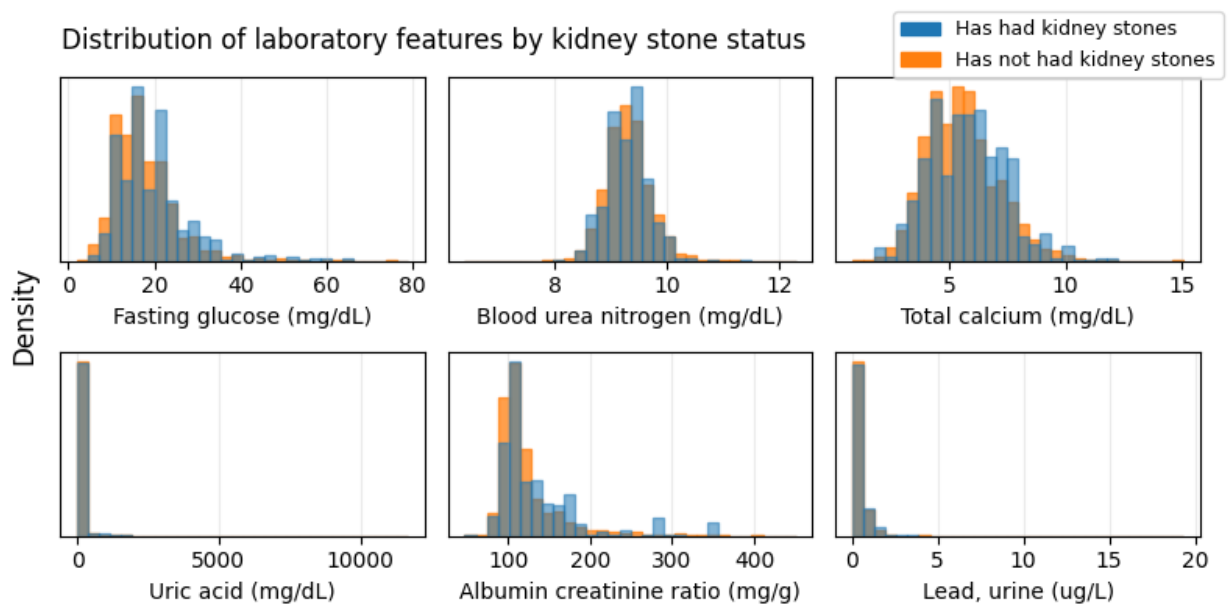


Figure 4: Laboratory Distribution

[Detailed description and interpretation of the laboratory distribution figure]

## 5.3 Dietary Analysis

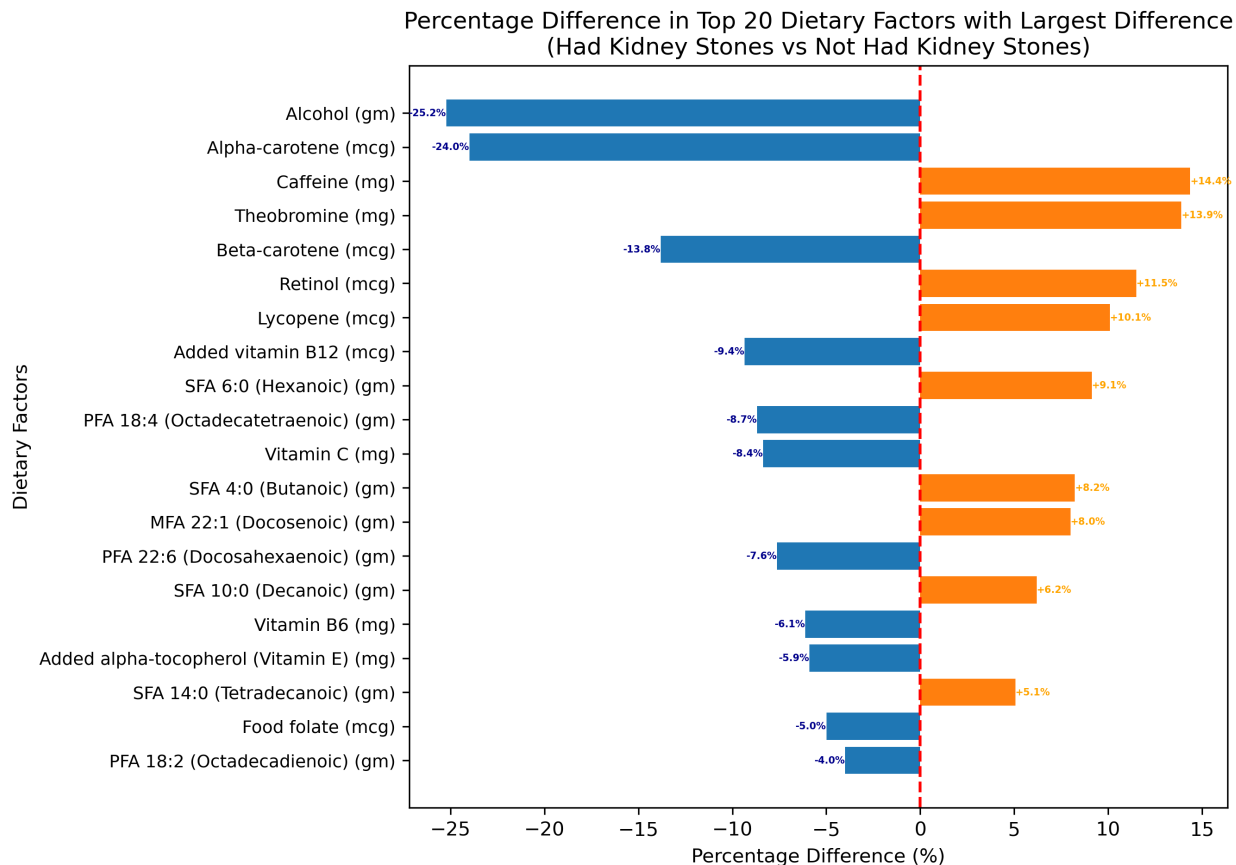


Figure 5: Dietary Differences

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

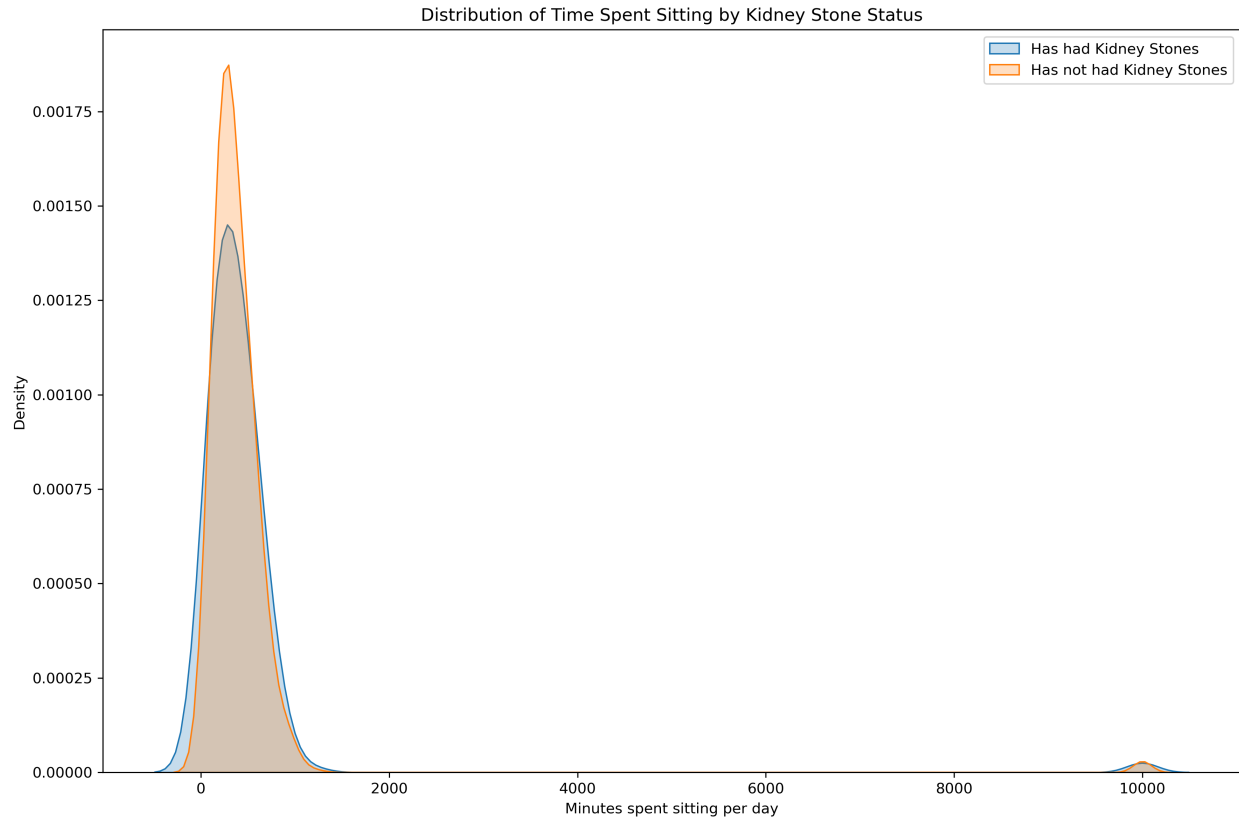


Figure 6: Time Spent Sitting

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.

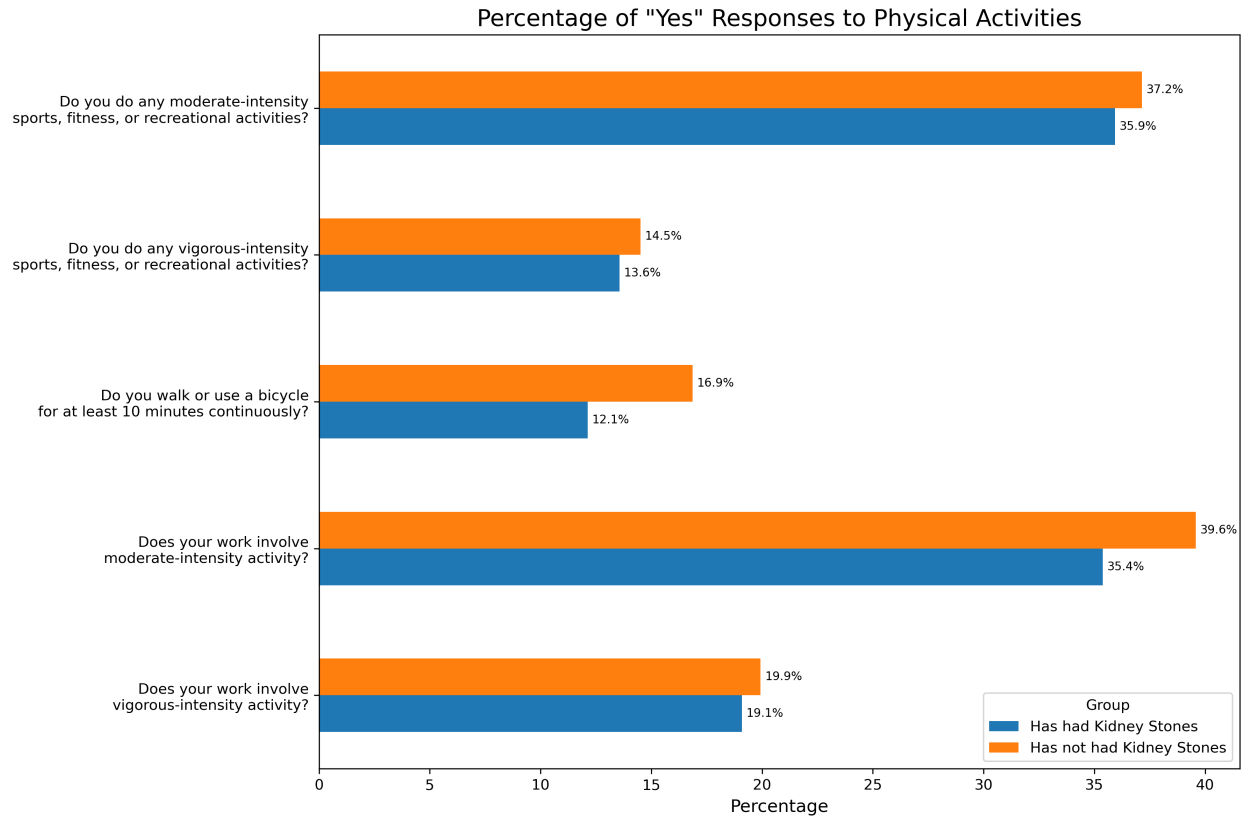


Figure 7: Physical Activities

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

These findings might 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.



proportion of Kidney Stones for people smoke less than 100 vs 100+ Proportion of Kidney Stones by Trouble of Sleep

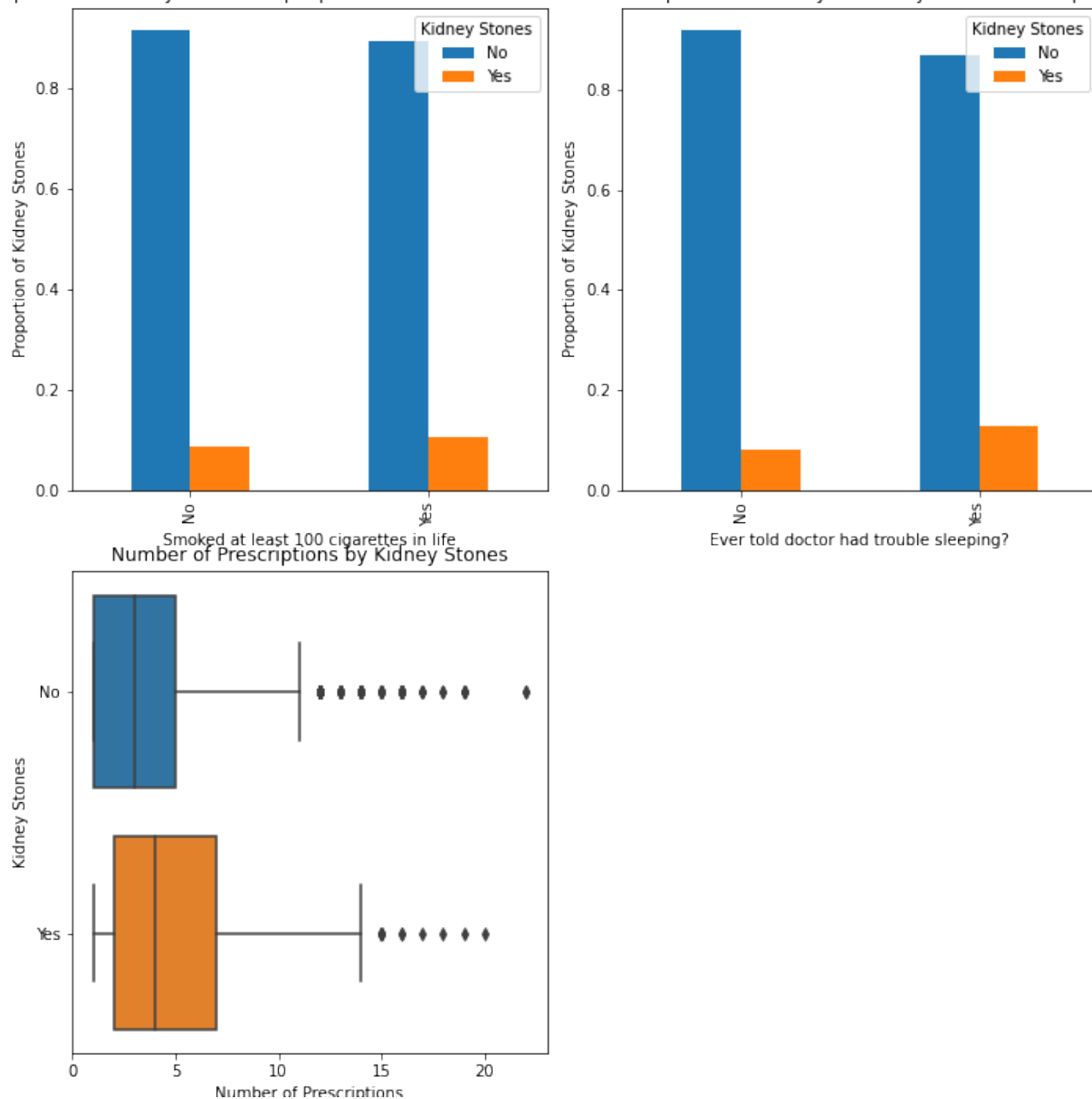


Figure 8: The risk of smoke to kidney stones. In this graph, you can see the proportion of people who have kidney stones or not for people with certain characteristics(smoke more than 100 times in their life, smoke less than 100 times in their life, don't know and refuse to answer). In this case, I will not account for people who answer I don't know and refuse. Since the number of them is low and I am not sure if I should assume that they smoke too little, so they think it is not important, or they smoke a lot. So just look at people who smoke more than 100 times in their life(Yes) and smoke less than 100 times in their life(No). You can see the proportion of kidney stones for people who answer yes is higher than those who answer no. This shows the correlation between the risk of kidney stones and smoking and it backs up the research that says there is a causation between the risk of kidney stones and smoking. So it means that smoking will increase the risk of developing kidney stones.

Figure 9: The risk of trouble sleeping to kidney stone Note(told the doctor about having trouble sleeping means that they have an issue about sleeping and didn't tell the doctor about having trouble sleeping means that they have don't issue about sleeping) In this graph you can see the proportion of people who have kidney stone or not for people with some certain characteristics(having a sleeping issue, don't have a sleeping issue,

don't know and refusing to answer). In this case, I will not account for people who answer I don't know and refuse. Since the number of them is low and I am not sure if I should assume they having trouble sleeping or not. So just look at people who having trouble sleeping(Yes) and people who aren't having trouble sleeping(No). You can see the proportion of kidney stones for people who answer yes is higher than those who answer no. This shows the correlation between the risk of kidney stones and smoking and it backs up the research that says there is a causation between the risk of kidney stones and having sleeping issues. So it means that having an issue sleeping(Short sleep time) will increase the risk of developing kidney stones.

Figure 10: The risk of the number of prescriptions taken for kidney stones In this graph you can see that I created a plot box to see on average the number of prescriptions taken by people who have kidney stones or not. It is shown that the number of people who have kidney stones tends to have a higher number of prescriptions taken compared to people who don't have kidney stones. From this, it shows the correlation between the risk of kidney stones and the number of prescriptions taken and it backs up the research that says there is a causation between the risk of kidney stones and the number of prescriptions taken. Since some prescriptions increase the risk of developing kidney stones. Also as the number of prescriptions taken increases, it is more likely for that person to have kidney stones. Since kidney stones are related to many medical conditions.

## 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]  
 Tinh Nguyen (300611544) - Risks Factor Exploratory Data Analysis (EDA), Identification of trends and patterns in the data, analysis of key variables(fig8,9,10)

## 9. References

[List references using a consistent citation style]

### **Smoke with increase the risk to develop kidney stones**

Xu, J., Deng, S., Song, L., Wang, W., Liu, Y., Li, Q., ... & Wu, W. (2023). The impact of secondhand smoke on the development of kidney stone disease is not inferior to that of smoking: A longitudinal cohort study. *BMC Public Health*, 23, Article 16116.

### **Short sleep time(sleeping issue) increase the risk to develop kidney stones**

Yin, Y., Zhang, Z., Zhang, X., Li, H., & Guo, W. (2022). Characterizing and assessing municipal solid waste management models in Chinese megacities: A case study of Shenzhen. *Sustainable Cities and Society*, 84, 104096.

### **Some prescription increase the risk of kidney stone**

NHS. (n.d.). Kidney stones - Causes. NHS UK.

### **Kidney stones are related to many medical conditions**

UF Health Urology. (n.d.). Stone disease. University of Florida, Department of Urology.