

# 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
5.5 Risk factor Analysis . . . . .	9
<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 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.

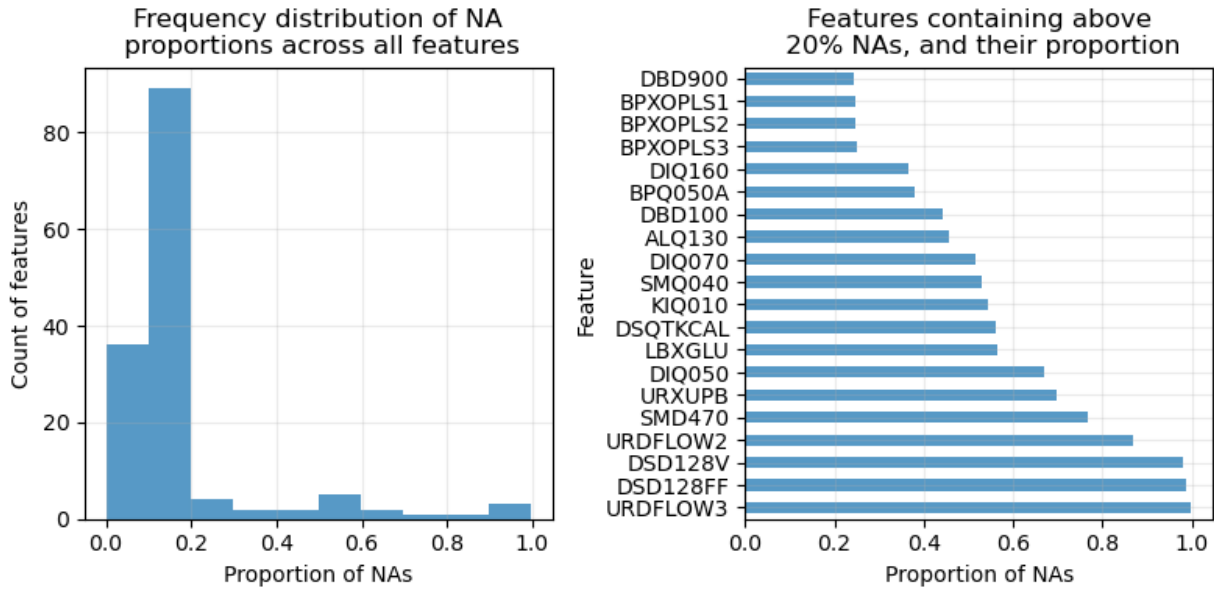


Figure 1: Distribution of missing data across features

### 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

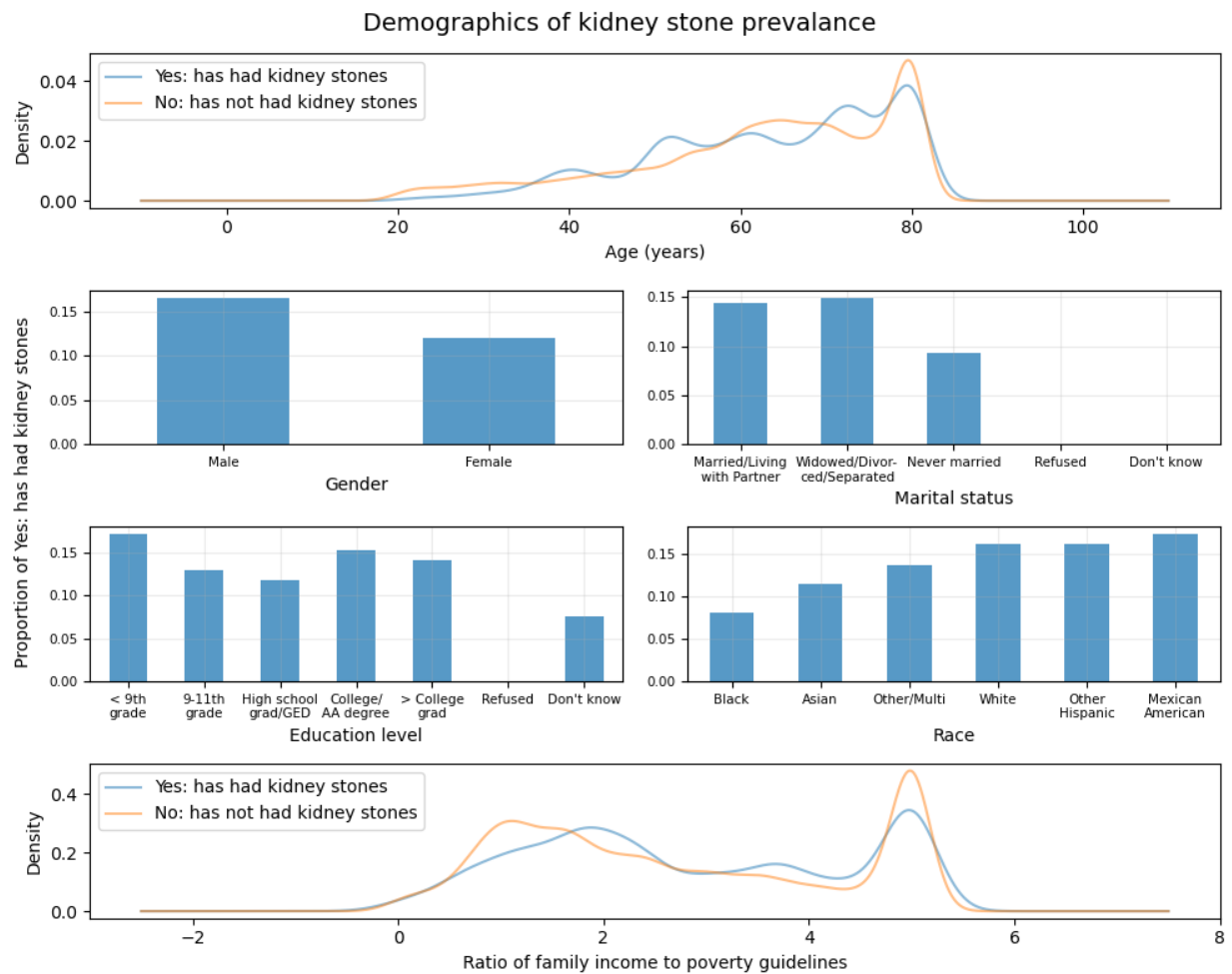


Figure 2: Demographic Kidney Stone Prevalence

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

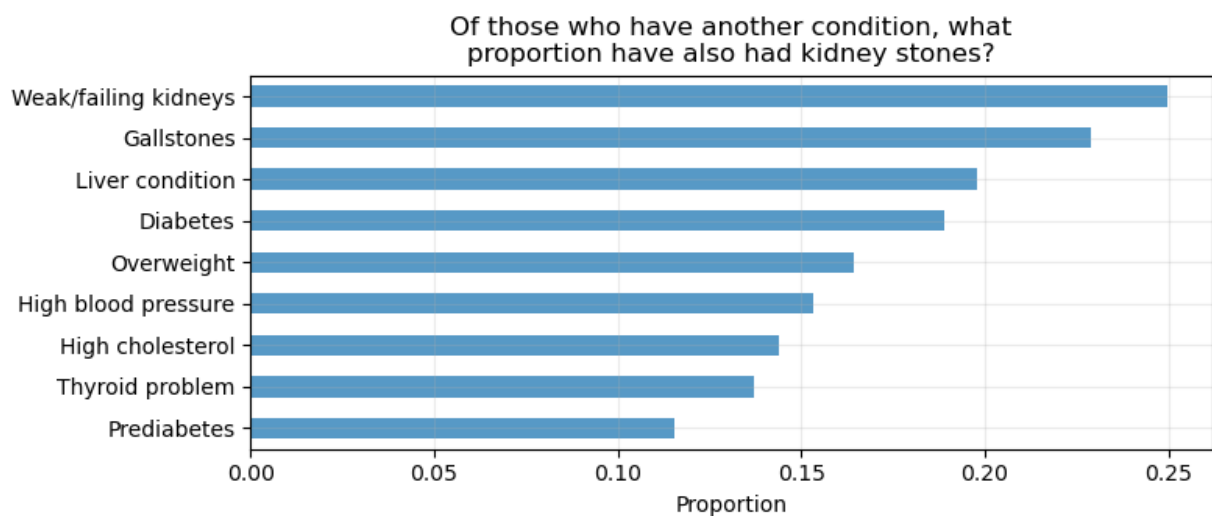


Figure 3: Condition Kidney Stone Proportion

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

## 5.2 Laboratory Analysis

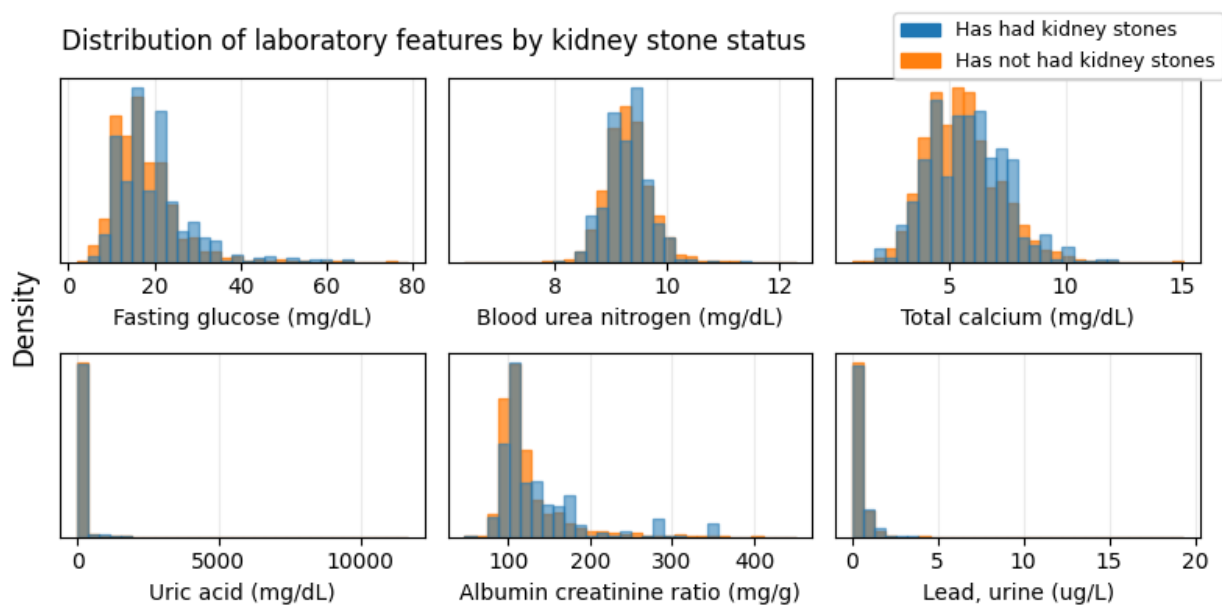


Figure 4: Laboratory Distribution

[Detailed description and interpretation of the laboratory distribution figure]

## 5.3 Dietary Analysis

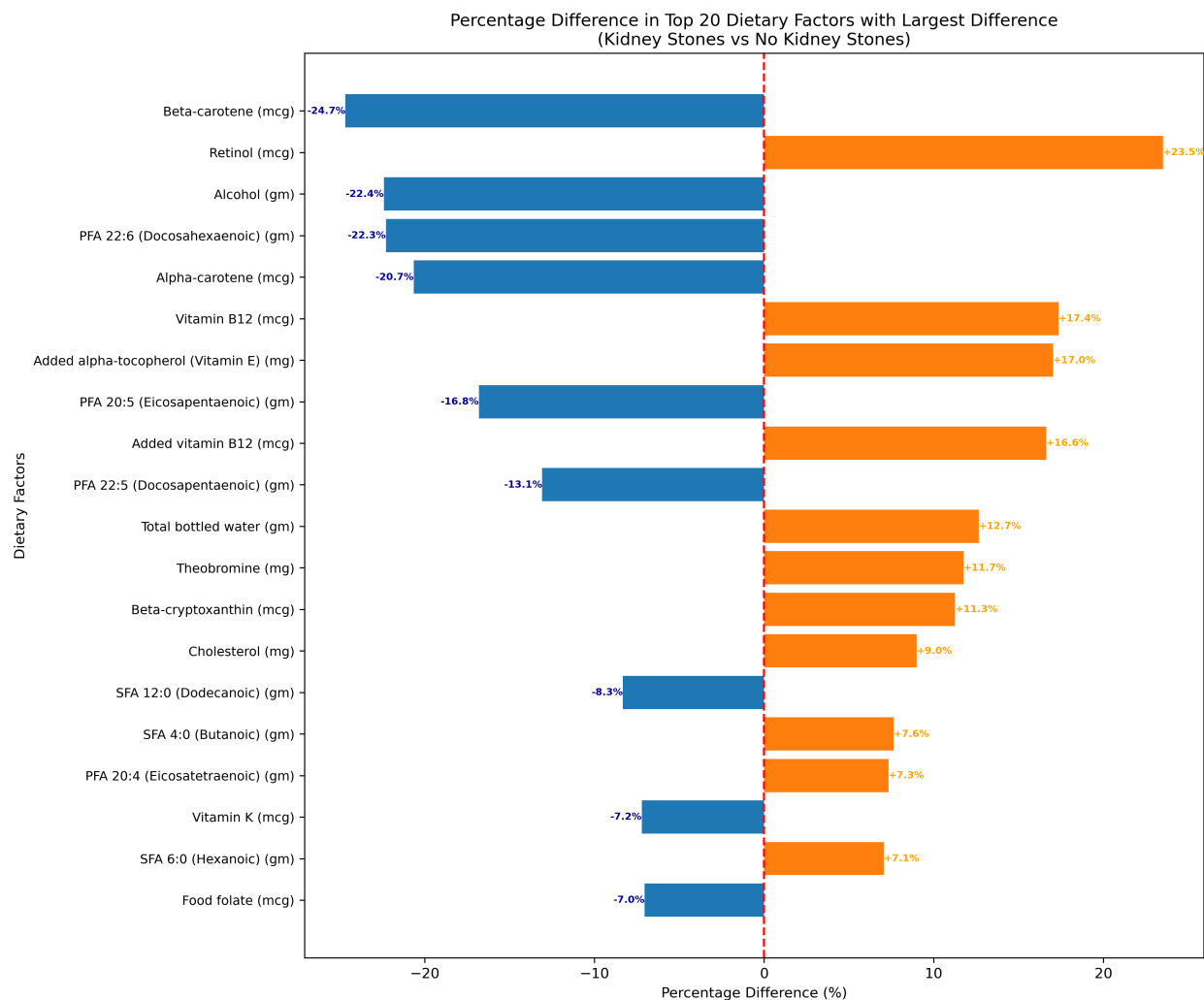


Figure 5: Dietary Differences

Retinol (a form of vitamin A) shows the largest positive difference, with individuals with kidney stones consuming approximately 23.5% more than those without. This suggests a potential positive association between retinol intake and kidney stone formation.

Conversely, beta-carotene (another form of vitamin A) displays the most substantial negative difference, with kidney stone formers 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 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 play a role in kidney stone formation.

Interestingly, alcohol consumption shows a large negative difference (-22.4%), indicating that individuals with kidney stones tend to consume significantly less alcohol. This finding challenges some traditional assumptions about alcohol and kidney stone risk.

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 prone to kidney stones and those who are not.

## 5.4 Physical Activity Analysis

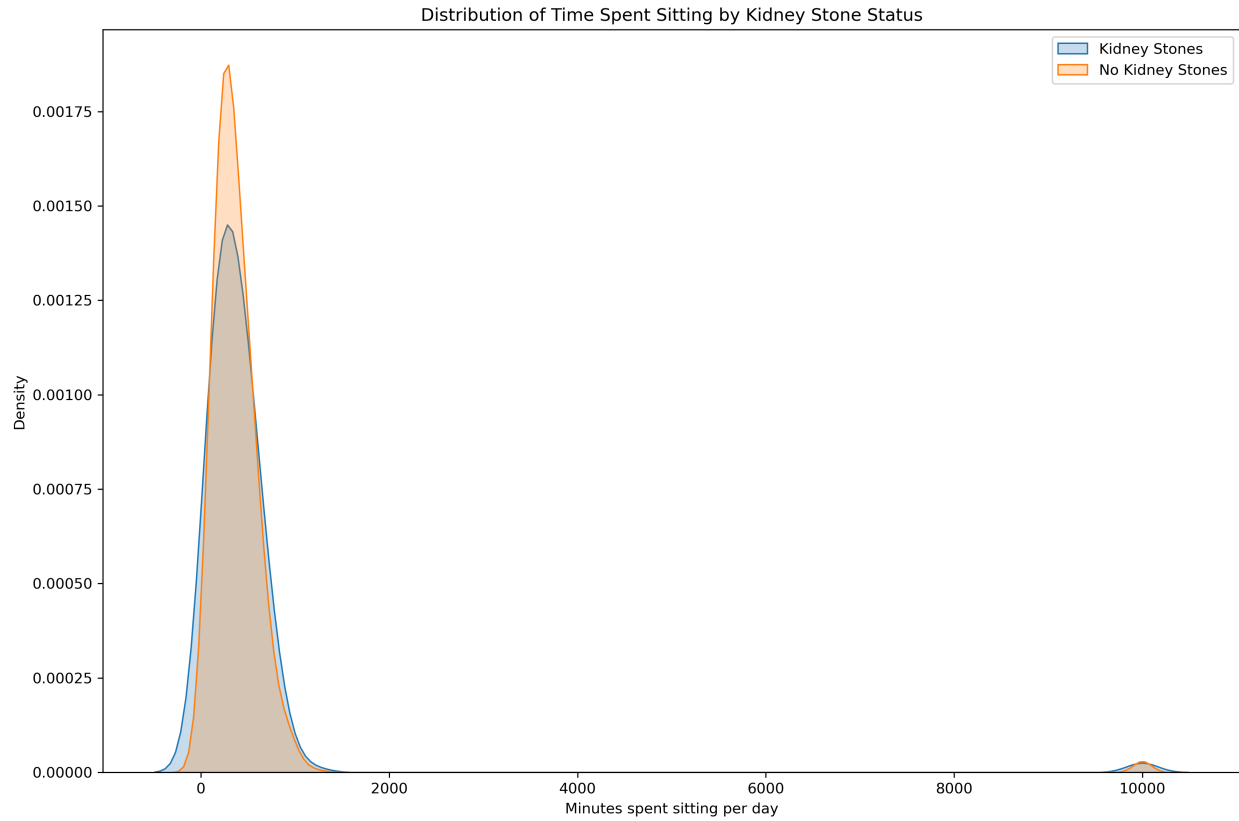


Figure 6: Time Spent Sitting

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.

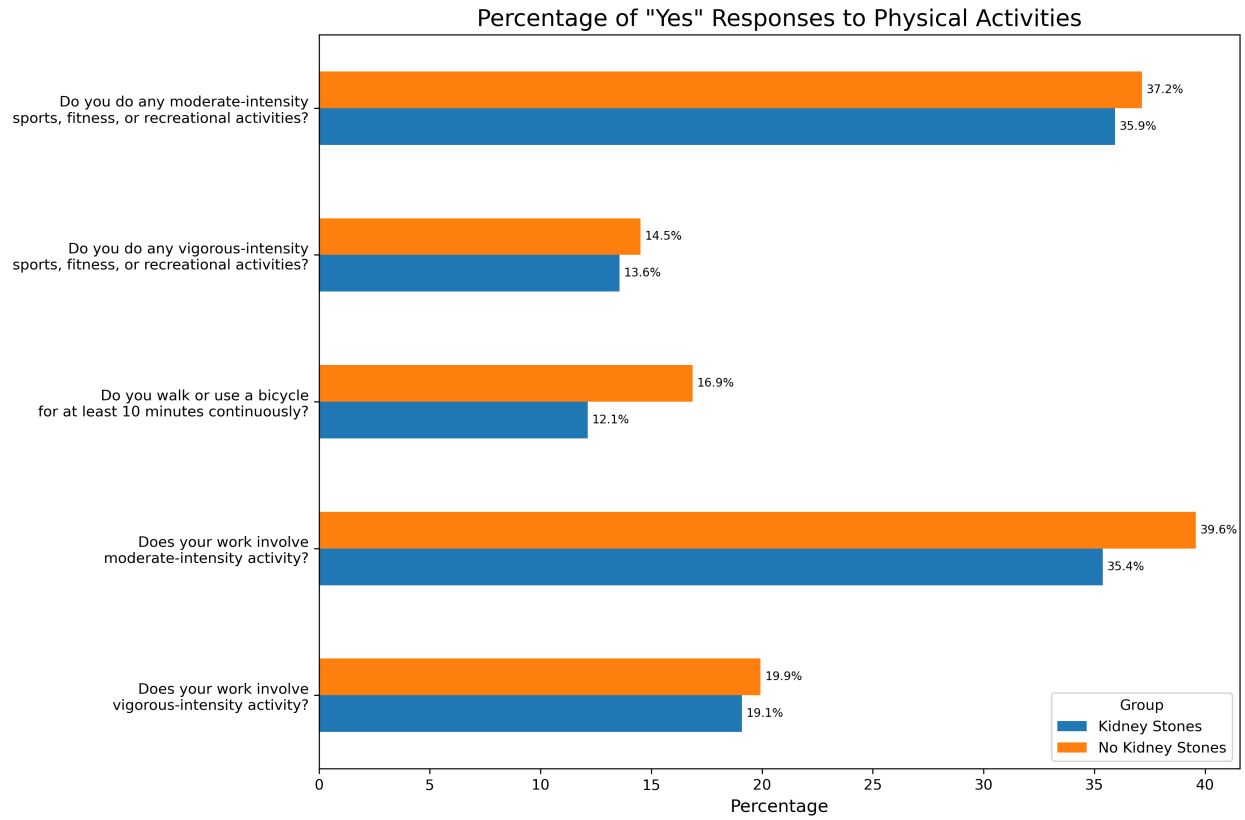


Figure 7: Physical Activities

[Add detailed description and interpretation of the physical activities figure]



## 5.5 Risk factor Analysis

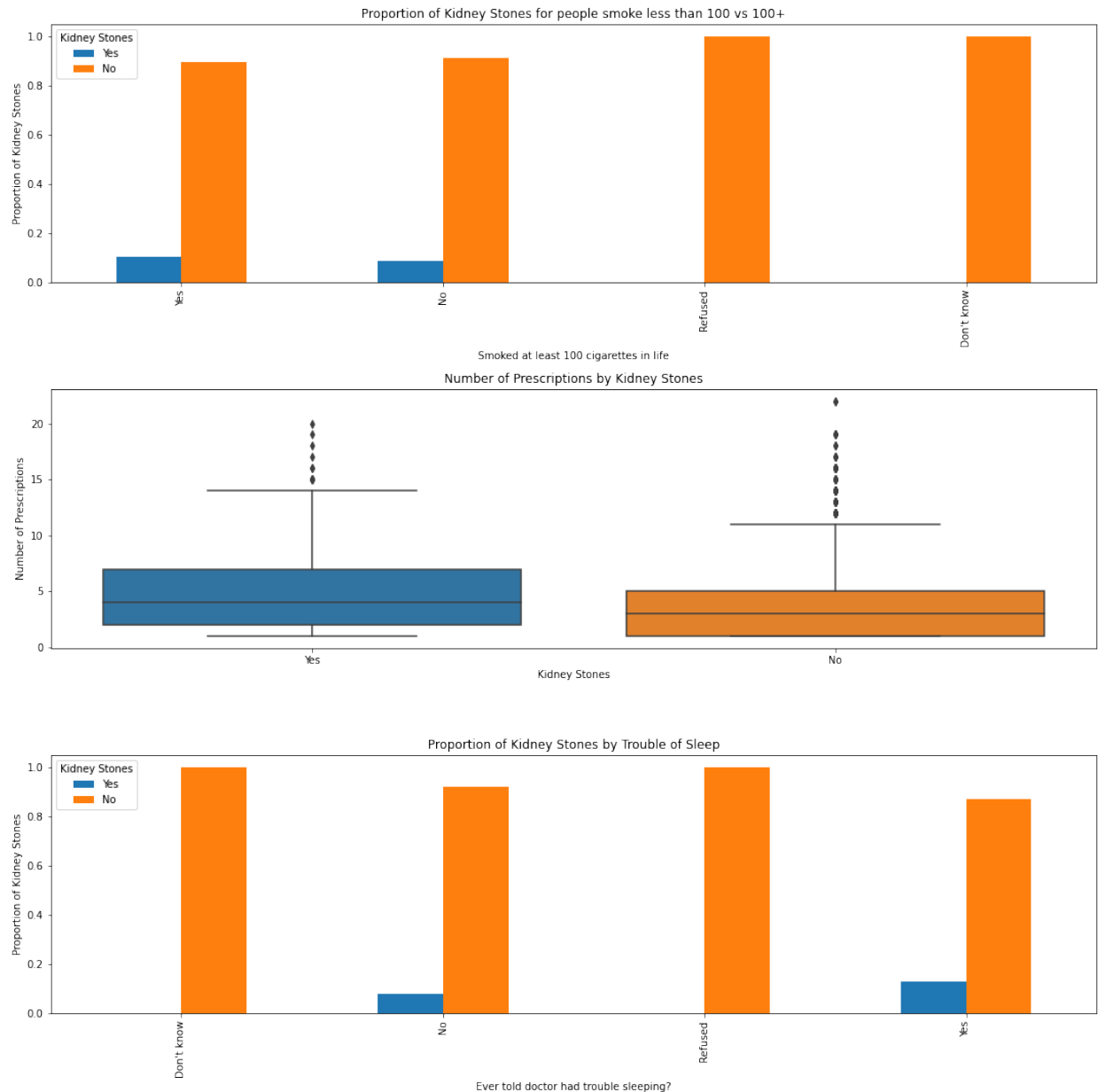


Figure 8: The risk of smoke to kidney stone In this graph you can see the proportion of people who have kidney stone or not for people with some certain characteristics(smoke more then 1000 time in their life,smoke less then 1000 time in their life, don't know and refuse to answer). In this case, I will not account for people who answer I don't know and refuse. Since 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 people who smoke more then 1000 time in their life(Yes) and smoke less then 1000 time in their life(No). You can see the proportion of kidney stone for people who answer yes is higher then who answer no. From this it show the correlation between risk of kidney stone and smoking and it back up the research that say their is causation between risk of kidney stone and smoking.

Figure 9: The risk of number of prescriptions taken to kidney stone In this graph you can see that I create a plot box to see in average the number of prescriptions taken by people who have kidney stone or not. It is show that the number of people who have kidney stone tend to have higher number of prescriptions taken

compare to people who don't have kidney stone. From this it show the correlation between risk of kidney stone and number of prescriptions taken and it back up the research that say their is causation between risk of kidney stone and number of prescriptions taken.

Figure 10: The risk of trouble sleeping to kidney stone Note(told doctor about having trouble sleeping mean that they have issue about sleeping and didn't told doctor about having trouble sleeping mean 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 number of them is low, and I am not sure if I should assume they they having trouble sleeping or not. So just look people who having trouble sleeping(Yes) and people who didn't having trouble sleeping(No). You can see the proportion of kidney stone for people who answer yes is higher then who answer no. From this it show the correlation between risk of kidney stone and smoking and it back up the research that say their is causation between risk of kidney stone and having sleeping issue.

## 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]