

### **Pre-Analysis Steps**

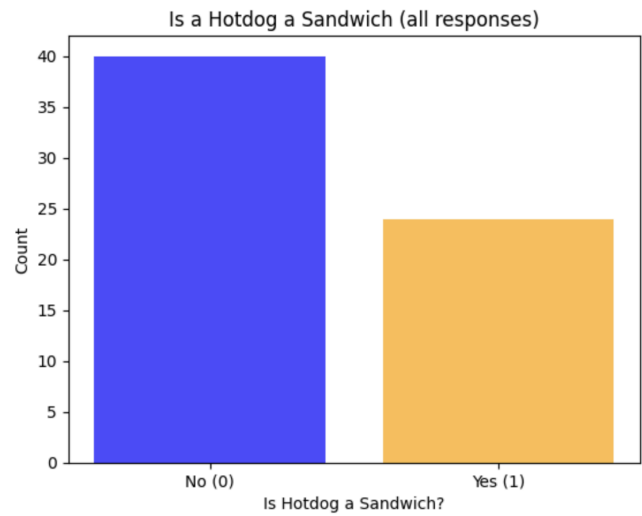
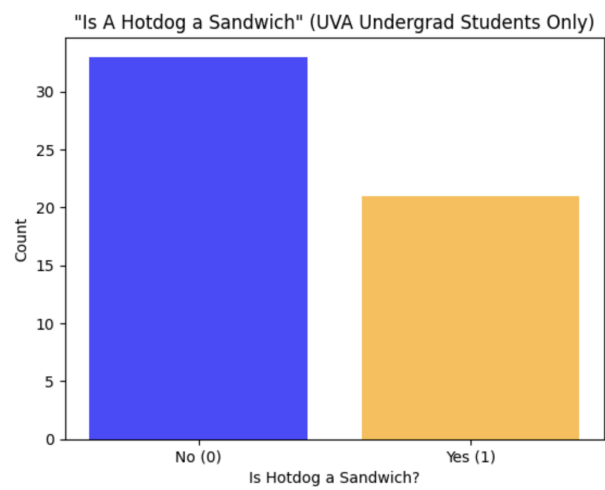
Before analyzing our dataset, which consisted of 9 columns and 64 entries, we performed several preprocessing steps to ensure clarity and usability. We cleaned the data by removing unnecessary whitespace and non-applicable values. To standardize the "academic year" column, we converted entries such as "First year" to "1" and "Fourth year" to "4." We excluded the timestamp and major columns, as they were not relevant to our analysis. Additionally, we simplified binary yes/no responses in the "Are you an undergraduate?" and "Do you think a hotdog is a sandwich?" columns by encoding them as 1 for "Yes" and 0 for "No." These steps helped streamline the dataset for effective analysis.

### **Analysis Methods**

We mainly used Python notebooks along with RStudio to analyze the data. After cleaning the dataset, we first proceeded to find the number of people who agreed that a hot dog was a sandwich vs. did not agree. In order to do this, we had to clean this column by transforming it into a true/false vector, with any answer beginning with "Y" or "y" being TRUE, and any answer otherwise being FALSE. After finding this (and creating a bar chart in the process), we then went on to create bar charts that separated the data by year (which we had to further filter).

### **Evaluation of Success**

To evaluate the success of our analyses we will look for any type of correlation between selected variables. We will look for positive and negative correlation among years for undergraduate students enrolled at the University of Virginia and will search for biases among binary variables such as gender or student status. We found a strong positive trend favoring a "Yes" to the hotdog question when increasing the year of participating students.



<Figure size 640x480 with 0 Axes>

