

Hot Dog Data Dictionary

Column	Description	Potential Response
timestamp	Indicates the day, month, year, hour, minute, and second an individual responded to the survey	month/day/year, hour:minute:second
hotdog.is.sandwich	Provides a simple yes or no answer to our research question: is a hot dog a sandwich?	"Yes" indicates that the user believes a hot dog is a sandwich. "No" indicates that the user does not believe a hot dog is a sandwich.
reason	Allows the user to explain the reasoning to their answer in the previous question	A string which contains an explanation behind the user's answer to the question "Do you believe a hot dog is a sandwich?". Free response so content will vary greatly response to response.
ungrad.or.grad	Answers the question: is the respondent enrolled at UVA as a graduate student or as an undergraduate student? This question presumes that the respondent is enrolled at UVA in some capacity.	"Undergraduate" indicates the respondent is enrolled as an undergraduate student at UVA, "Graduate" indicates the respondent is enrolled as a graduate student as UVA.
major	This question gives further information on the respondent regarding what they are studying, presumably at UVA	A string which contains a valid field of study/major.
year	This question gives further information on the respondent regarding what their current academic year at UVA is, should they identify as an undergraduate student	No response indicates that the user is a graduate student. "First year" indicates a first year student, "second year" indicates a second year student, "third year" indicates a third year student, and "fourth year" indicates a fourth year student.
state.or.country	This question gives further information on the respondent, regarding the state or country they were born in.	A string which contains a valid state in the United States or country. Answers will vary as this is a free response question.

--	--	--

Md format:

Hot Dog Data Dictionary

****Column****	****Description****	****Potential Response****
timestamp	Indicates the day, month, year, hour, minute, and second an individual responded to the survey	month/day/year, hour:minute:second
hotdog.is.sandwich	Provides a simple yes or no answer to our research question: is a hot dog a sandwich?	"Yes" indicates that the user believes a hot dog is a sandwich. "No" indicates that the user does not believe a hot dog is a sandwich.
reason	Allows the user to explain the reasoning to their answer in the previous question	A string which contains an explanation behind the user's answer to the question "Do you believe a hot dog is a sandwich?". Free response so content will vary greatly response to response.
ungrad.or.grad	Answers the question: is the respondent enrolled at UVA as a graduate student or as an undergraduate student? This question presumes that the respondent is enrolled at UVA in some capacity.	"Undergraduate" indicates the respondent is enrolled as an undergraduate student at UVA, "Graduate" indicates the respondent is enrolled as a graduate student as UVA.
major	This question gives further information on the respondent regarding what they are studying, presumably at UVA	A string which contains a valid field of study/major.
year	This question gives further information on the respondent regarding what their current academic year at UVA is, should they identify as an undergraduate student	No response indicates that the user is a graduate student. "First year" indicates a first year student, "second year" indicates a second year student, "third year" indicates a third year student, and "fourth year" indicates a fourth year student.
state.or.country	This question gives further information on the respondent, regarding the state or country they were born in.	A string which contains a valid state in the United States or country. Answers will vary as this is a free response question.

Task 2: Register and Conduct Analysis

Before analyzing the data, we cleaned the data by making all survey entries lowercase. We also renamed some of the column names so that they would be easier to work with in R. For the data column regarding majors, we used R code to change any "cs" responses to "computer science" since we had many computer science majors who answered the survey differently. For the other majors, we went back into the excel file and manually changed responses with multiple majors to "double major." We did the same for the data column regarding state/country of residence. Since our data set was small enough that manually changing entries would not take much time, we went back into the excel file and changed state names so that they were referred to using consistent language. We then re-uploaded the excel file into R so that we could use the newly cleaned dataset for analysis.

Our exploratory data analysis consisted of displaying the cleaned data in visual form, primarily in the use of bar graphs. We examined possible correlations between our research question (is a hot dog a sandwich?) to respondents' answers and other information about them. We first examined the correlation between our research question and the general student-respondent body (both undergraduate and graduate students) which revealed that most believe a hotdog is not a sandwich (62 voted no and 30 voted yes). Then, we examined the correlation between our research question and undergraduate respondents' academic year, finding that the highest number of students who did *not* think hotdogs are sandwiches are fourth year students. At the same time, the highest number of students who *do* think hotdogs are sandwiches are also fourth years- though more fourth year students believe that hotdogs are not sandwiches. Per student year, the majority of students believe that a hotdog is not a sandwich. More specifically, across the undergraduate student-respondent body, our analysis revealed that the highest number of respondents were computer science majors.

When examining the data, we found a potential bias in the disparate proportions of graduate versus undergraduate students, with the vast majority of respondents identifying as undergraduates. Since this course primarily consists of upperclassmen, we also recognize that a majority of the respondents are either third or fourth years. There are a total of 91 respondents with 2 first year responses, 7 second year responses, 23 third year responses, and 52 fourth year responses. In conclusion, we found that from the data collected the consensus is that the majority of UVA students do not consider a hot dog to be a sandwich. However, we do take this conclusion with a grain of salt, as we do not consider our sample data representative of the entire student population.

Our null hypothesis for statistical analysis is that the number of students who answered "Yes" for a hot dog being a sandwich is equal to the number of students who answered "No." Our alternative hypothesis is that the number of "Yes" responses we received is not equal to the number of "No" responses. To test this hypothesis, we used a one sample proportion test where we set the probability, p , equal to 0.5 to theoretically test the odds that we receive an exact 50/50 split in terms of responses. With a sample size of 92 responses, where we received 30 "Yes" responses coded as a success, our test presented a p-value of 0.001111. Since our p-value is less than the alpha value of 0.05, we reject our null hypothesis in favor of our alternative hypothesis and conclude that the number of students who believe a hotdog is a sandwich is not equal to the number of students who do not believe a hotdog is a sandwich.

```
binom.test(x = 30, n = 92, p = 0.5, alternative = "two.sided")
```

Exact binomial test

```
data: 30 and 92
number of successes = 30, number of trials = 92, p-value = 0.001111
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.2319638 0.4317606
sample estimates:
probability of success
      0.326087
```

```
prop.test(x = 30, n = 92, p = 0.5, alternative = "two.sided", correct = TRUE)
```

1-sample proportions test with continuity correction

data: 30 out of 92, null probability 0.5
X-squared = 10.446, df = 1, p-value = 0.001229
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
0.2341242 0.4327640
sample estimates:
p
0.326087

GRAPHS:





