

How much are we sleeping? What are we eating?

Week 1, Lecture 2

<http://live.ds306.org>

September 3, 2026

Announcements

- Submit the tutorial survey by **tonight**.
- PS1 is live, due next Wednesday.
- I don't manage enrollment or the waitlist. Email statsugradprogram@umich.edu.

Lecture 1 recap

```
library(tidyverse)
theme_set(theme_minimal(base_size = 13, base_family = "sans"))

survey <- read_csv(
  "class_survey.csv",
  show_col_types = FALSE
)
```

In Lecture 1, we:

- asked for the median reported height;
- discovered values recorded in meters, inches, and centimeters;
- worked through the height standardization using `if_else()`; and
- briefly used `filter()`, `mutate()`, and `|>`.

Today we will go over these concepts slowly and in more detail, and use them to answer new questions.

Today's questions

1. Did at least half of the respondents report eight or more hours of sleep last night?
2. Do food preferences vary based on continent of birth?

Question 1: Did at least half sleep eight hours?

Before seeing the sleep responses, predict whether at least half of the respondents reported eight or more hours of sleep last night. Estimate the percentage and give one reason for your prediction.

Keep the columns we need

- Make life simpler by isolating only the columns of data we need to answer the question.
- `select ( )` retains columns from a data frame. It does not remove rows.

```
sleep_data <- select(  
  survey,  
  response_id,  
  sleep_value,  
  sleep_unit  
)  
  
glimpse(sleep_data)
```

Distribution of responses

`summary(x)` summarizes a vector `x`.

```
summary(sleep_data$sleep_value)
```

What do you learn by summarizing `sleep_value`?

Responses are in two units

```
table(sleep_data$sleep_unit)
```

The survey contains 94 responses in hours and 4 in minutes.



Values recorded in different units must be converted to a common unit before they can be compared.

Plan for answering question 1

1. Convert all responses to hours.
2. Determine which respondents reported eight or more hours of sleep.
3. Count the number of respondents who reported eight or more hours.
4. Compare that to the total number of respondents.

Create new columns with mutate()

mutate(data, ...) creates or changes columns.

```
df <- mutate(  
  sleep_data,  
  sleep_hours = if_else(  
    sleep_unit == "minutes",  
    sleep_value / 60,  
    sleep_value  
  )  
)
```

This accomplishes step 1 of our plan.

Create a new column for eight-hour sleep

```
df <- mutate(df, got_eight_hours = sleep_hours >= 8)
```

This accomplishes step 2 of our plan.

Reintroducing `filter()`

`filter()` retains the rows for which a condition is TRUE.

```
eight_hour_sleep <- filter(df, got_eight_hours)
```

This keeps the rows we need for step 3. We still need to count them.

Exercise

Combine `mutate` and `filter` to determine how many hours survey respondents R016 and R031 slept.

```
# combine mutate and filter
```

Building a check table



After manipulating data, it's good to manually inspect some results to make sure the calculation did what you intended.

```
slice_sample(df, n = 5)
```

Calculate the final answer

Putting everything together:

```
# count the respondents who slept eight or more hours
```

Starting from the data frame we just created, explain what additional steps you need to take to answer the question, “Did at least half of the respondents sleep eight or more hours?”

Change the cutoff

What percentage of survey respondents slept at least nine hours?

What percentage slept fewer than six hours?

The `|>` operator

- In answering the previous questions, the operations appeared in the opposite order from their execution.
- To make our code more readable, we have the `|>` (“pipe”) operator.
- `x |> f()` means `f(x)`: pass `x` as the first argument to `f()`.

Example

```
eight_hour_sleep <- survey |>
  select(response_id, sleep_value, sleep_unit) |>
  mutate(
    sleep_hours = if_else(
      sleep_unit == "minutes",
      sleep_value / 60,
      sleep_value
    )
  ) |>
  filter(sleep_hours >= 8)
```

Food preferences and continent of birth

Let's study `favorite_food` and `birth_continent`, beginning with each variable separately.

Predict before looking

In Lecture 1, we predicted the class's median height before inspecting the data. Make two new predictions before running any code.

Among the 98 survey responses:

1. Which favorite food do you predict was selected most often? Why?
2. Which continent of birth do you predict was reported most often? Why?

Food preferences with `table()`

`table()` counts each distinct value in a vector.

```
table(survey$favorite_food)
```



This is a **distribution** of the categorical variable `favorite_food`.

Food preferences with `geom_bar()`

A bar plot is another way to visualize the distribution of a categorical variable:

```
ggplot(survey, aes(y = favorite_food)) + geom_bar()
```

Continent of birth with `table()`

What percentage of survey respondents were born in Asia? North America? Elsewhere? (Hint: use the `table()` function and divide by the total number of respondents.)

Continent of birth with `geom_bar()`

```
ggplot(survey, aes(y = birth_continent)) +  
  geom_bar()
```

The bar graph makes imbalances between the categories more visible.

What can't these two graphs tell us?

One graph shows food choices for the whole class. The other shows continents of birth for the whole class. Using both these graphs, can we tell whether food choices differed between respondents born in Asia and respondents born in North America? Why or why not?

Did favorite-food choices differ between the groups?



Two one-variable distributions cannot tell us whether the variables are related. We must examine the variables jointly.

We will calculate food preferences separately for respondents born in Asia and respondents born in North America.

Build a two-way table

```
food_counts <- table(  
  survey$birth_continent,  
  survey$favorite_food  
)
```

```
food_counts
```

Rows are continent groups, columns are food choices, and each cell is the number of responses.

What stands out in the counts?

Which foods appear most common in each birth-continent group?
Describe one similarity and one difference between the two rows of counts.

Comparing the two groups

Ten respondents reporting Asia chose sushi. Twenty-six of the 80 respondents reporting North America chose sushi. Does this mean sushi is more popular in North America? Explain.

Normalization

- To compare the groups, divide every count by the total for its row.
- The result is two rows of percentages, each summing to 100%.

```
prop.table(food_counts, margin = 1) * 100
```

Predict the within-group percentages

Based on the percentages calculated above, is sushi more popular with Asian or North American students?

Dividing counts by column totals

```
prop.table(food_counts, margin = 2) * 100
```

- `margin = 2` divides every cell by the total of its column. Each column therefore sums to 100%.

Joint and conditional distributions

The count of 10 respondents who reported Asia and chose sushi answers three different questions, depending on the denominator:

- $10 / 98 = 10.2\%$ reported Asia **and** chose sushi among all respondents. This is a *joint distribution*.
- $10 / 18 = 55.6\%$ chose sushi **among respondents who reported Asia**. This is a *conditional distribution* of food preference given continent of birth.
- $10 / 36 = 27.8\%$ reported Asia **among respondents who chose sushi**. This is a *conditional distribution* of continent of birth given food preference.

Compare the two tables

Using the row percentages, describe one difference in food choices between the two continent groups. Using the column percentages, describe one difference in reported continent among people choosing different foods.

Do students in Asia and North America differ in sushi preference?

- 10 of 18 (55.6%) of respondents from Asia preferred sushi.
- 26 of 80 (32.5%) respondents from North America preferred sushi.
- Can we conclude that students in Asia and North America differ in their preference for sushi?

What sort of statistical test could we use to settle this question?

Simplifying the table for a two-proportion test

- The full table contains eight food choices.
- For one simple test, combine them into two outcomes.
- We want R to make a **2×2 table** with the following structure:

Reported continent	Chose sushi	Did not choose sushi
Asia	XX	XX
North America	XX	XX

Step 1: Create the two outcomes

```
sushi_data <- survey |>
  mutate(
    sushi_choice = if_else(
      favorite_food == "Sushi",
      "Chose sushi",
      "Did not choose sushi"
    )
  )
```

`mutate()` gives every row one of the two values in `sushi_choice`.

Step 2: Fill the table

```
sushi_table <- table(  
  sushi_data$birth_continent,  
  sushi_data$sushi_choice  
)
```

```
sushi_table
```

Could random variation explain the difference?

Suppose sushi were equally popular in the two groups. Could random samples of 18 and 80 people produce percentages of 55.6% and 32.5%, or would a gap that large be surprising? Make a prediction before we run the test.

Test the two proportions

The test of proportions measures how surprising a 23.1-point gap would be with group sizes of 18 and 80.

```
prop.test(sushi_table)
```

Interpret the result

What does the test show us? Interpret the p-value. Can we conclude that sushi is more popular in one group than the other?

Answer to the sushi question

- In this survey, 55.6% of respondents reporting Asia chose sushi, compared with 32.5% of respondents reporting North America.
- The difference was not statistically significant at the 5% level ($p = 0.118$).
- The difference is present in our survey, but the test does not provide strong evidence that it is more than random variation.
- This test concerns sushi only.



The observed difference describes these respondents. The test asks whether random variation could plausibly produce that difference; it does not make the voluntary survey representative.

Exit question

Suppose we want to repeat the 2×2 test for burrito versus not burrito.

1. What two outcomes form the columns of the table?
2. What numbers would you use for x and n in `prop.test(x, n)`?
3. What equal-proportions assumption would the test begin with?

Week 1, Lecture 2

Closed September 3, 2026 at 4:28 PM

16 class questions · 878 anonymous responses · 1 student questions

The following slides preserve what students submitted during class. No names or login information are included.

Before seeing the sleep responses, predict whether at least half of the respondents reported eight or more hours of sleep last night. Estimate the percentage and give one reason for your prediction.

97 anonymous responses

Predicts less than half (numeric ~20–40%)

54

Respondents who estimated well under 50% (commonly ~25–40%) and cite college workload, busy schedules, or typical poor sleep habits as the reason.

Predicts less than half (numeric ~40–49%)

17

Respondents who predict just under half (around 40–49%), often noting early semester schedules or mixed factors.

Other responses

11

Responses the model could not place reliably.

Predicts at least half (numeric ~55–70%)

9

Respondents who expect over 50% (typically 55–70%), reasoning that it's syllabus week/early semester so students get more sleep or follow recommended amounts.

Response themes for question 1

Predicts about half ($\approx 50\%$)

3

Respondents who predict roughly half or express uncertainty near 50% often because classes haven't fully started or schedules are variable.

Short/unclear numeric or single-word answers

3

Responses that are very short, unclear, or give no substantive reason.

Student responses

1. 30%, although it's the first week of school, it's rare that college students sleep more than 7 hours these days since people have activities outside of extracurriculars going on. The 30% accounts for the probability of students that simply keep healthy routines by estimation.
2. 30%, data science studies tend to prioritise studying over sleep, and it's also first week, so some people may not take a strict school routine as seriously.
3. 35%, I would say that I get 8hrs 35% of the time so id expect that to reflect in the population
4. around 25%
5. Before seeing the sleep responses, I do not think that at least half of respondents slept eight or more hours last night. Most students have early class times and/or stay up late doing school work or club meetings, so I'd imagine only about 25% got eight or more hours of sleep last night.
6. I believe that about 40% slept at least 8 hours. Although I think on a typical college night its probably around 25% who slept over 8 hours. I think this will be a little higher since it is earlier in the semester and there is a lot less work.

Student responses

1. I believe that less than half of the class sleeps eight hours my guess would be 34% because I think college and especially these majors are a good amount of work and students don't sleep that much
2. I believe that less than half of the students slept 8 hours because it is a school night, and many students may have had early classes on the Tuesday, my guess would be around 40% of students got 8+ hours.
3. I do not think at least half of respondents slept 8 hours or more. Using a smaller sample size of my friends, I would estimate that around 35-40% of use actually get 8 hours so my guess would be around 37-38% of respondents had 8 or more hours of sleep.
4. I do not think half of the respondents reported sleeping at least eight hours. I estimate the real percentage to hover around 30%. With busy lives and early classes, college students are most likely to not get eight hours a night.
5. I do not think more than half the respondents had more than 8 hours of sleep. I would guess 30% of people did because of the culture surrounding sleep deprivation in American college students

Student responses

1. I do not think that at least half the respondents got 8 hours of sleep, everyone in the class is a college student and most are in their late teens and 20s and in my experience only a handful get a full nights rest. I think like 30% is more realistic.
2. I don't think half of the respondents reported eight or more hours of sleep. I would estimate the percentage to be 35%, lower than half, given that this is a stats course it seems like workload for these students may be greater than average.
3. I don't think so. Personally I sleep 7 hrs per night, so I think probably only 40% of the students could reach the 8-hr sleep goal based on my own experience.
4. I don't think that half of the class got 8+ hours of sleep. I would guess around 35% got less than 8 hours because most college kid are very busy and stay up later.
5. I estimate less than half of students slept more than eight hours. College students tend to sleep less than the suggested amount. I would guess that closer to 30% slept more than eight hours.

Student responses

1. I expect about 40% of respondents to have reported sleeping eight or more hours. I think the average number of hours slept is somewhere around 7, so somewhat less than half might have slept for 8 or more.
2. I predict about 35% reported at least 8 hours because college students typically do not have good sleep schedules.
3. I predict that half of the reported students did not get at least 8 hours of sleep last night. If I had to give a percentage, I would say that around 40% of students that responded to the survey received at least 8 hours of sleep and this is less than half because there were all college students that filled out the survey and most college students tend to stay up late at night studying or being with friends
4. I predict that half of the students didn't get 8 hours of sleep. The one reason is that the survey was taken before classes started so students could have been flying in and moving in. (not enough time to sleep). Maybe 40% of the students got 8 hours of sleep

Student responses

1. I predict that half of the students do not get 8 or more hours of sleep a night. I think it is closer to around 30% of people since as college students we tend to stay up later
2. I predict that less than half of respondents slept more than 8 hours, but not significantly less than half, maybe around 40%. This is because it is the beginning of the semester and many students may be running on late-night summer sleep schedules, however since it is the beginning of the semester and there isn't significant schoolwork yet, many students are still sleeping decently.
3. I predict that less than half of the students slept 8 hours or more, around 40%. I predict this because college students are known for sleeping poorly.
4. I predict that more than half of the respondents did not report getting 8 or more hours of sleep. I think this because most students are getting around 6 hours with work, social media, and other factors

Student responses

1. I think 20 percent of the class has reported eight or more hours of sleep and the rest of the class reported below. I think this before the typically the average length of sleep is 7 hours and given external circumstances like college
2. I think 29% of students slept more than 8 hours. This is because I don't think most college students sleep 8 hours. Also, this class is mainly upperclassmen I think, and I think they might sleep less than average too.
3. I think 30% of students slept at least 8 hours because college students tend to sleep less than 8 hours historically on average.
4. I think about 40% slept eight or more hours because its near the start of the semester and people may be busy moving in.
5. I think it will be less than half of eight or more hours of sleep. Maybe around 40 percent. Given that this is a data science course and a majority of students are STEM and they have a lot of homework, I feel like they don't get that much sleep.

Student responses

1. I think less than half of the respondents reported eight or more hours of sleep, because the survey targets are college students, and college students aren't known for getting a lot of sleep.
2. I think less than twenty percent of respondents slept eight or more hours the night before the first day of class. Less than half of all adults sleep 8+ hours a night in general, yet these were special circumstances (welcome week, college-aged rather than just adults).
3. I think maybe 30% of students reported that they slept 8 or more hours. I think this because I feel like not a lot of people in college prioritize sleep. Also, its the first week of school so everyone is still figuring out their schedules.
4. I think that 35% of the respondents actually got 8hrs of sleep. Nobody sleeps that much.
5. I think that about 30% of responses said that they slept at least 8 hours. I feel like in college, students tend to not get as much sleep as they intend to because of busy schedules and other activities.

Student responses

1. I think that less than half got 8 hours or more, around 35 percent. This is because college students are not great with getting a healthy amount of sleep.
2. I would assume less than 50% of respondents reported 8 hours or more of sleep. I would guess about 30% due to college students likely getting less sleep on average.
3. I would guess that about 40% of the class slept eight or more hours last night. With it being the first week of school, a lot is going on right now, so I doubt that more than half the class slept eight or more hours. However, because there aren't any exams yet, I doubt that this percentage is super low.
4. I would guess that less than half of the respondents got 8 or more hours of sleep the night before the survey maybe around 40%, since it is the first week of the semester and peoples schedules are changing and people are staying up later.
5. I would predict that around 30% of respondents reported getting more than 8 hours of sleep, this is because among college students I have found that a majority get less than 8 hours

Student responses

1. I would say about a third of the class slept 8 or more hours. College students are busy with school and work and so don't get much sleep at all. But it's also true that many people probably filled out this survey on the Sunday before classes, when they were able to get more sleep.
2. Less than half of all students slept for eight or more hours, probably in the range of 30%. College is very busy and 8 hours is seen as the healthy range you can compromise with.
3. Less than half will report 8 or more hours; it's college.
4. Maybe 20% of students got 8 hours of sleep. The distribution follows a normal distribution most likely centered around 7 hours of sleep or the recommended minimum amount of sleep for adults. The college student population is probably left tailed with more observations below 7 hours.
5. Maybe 25% slept 8 hours or more, most college students have too much work/activities to get 8 hours, and it's the first week of class
6. My estimation is that under half of 50% of the students who responded to the survey sleep at least 8 hours per night. I estimate 39% sleep over 8 hours.

Student responses

1. no I do not think half of us slept 8 hours last night.
2. No I don't think half of students slept 8 hours, I think 43% of students slept 8 hours just based off of how college students usually sleep
3. No I thin it is less than half. I would say 25% got 8 or more hours of sleep last night.
4. No, I don't think more than half sleep 8 hours. I would predict about 30/35% of people sleep 8 hours or more, because the average student typically stays up late at night and has to get up relatively early for classes
5. No, I think it is around 28% of people and this is because many college students choose to prioritize other acitivites over sleep
6. No, I think stress makes people sleep less on average, so I think people are sleeping a little less for the first week of classes. I think about 40% of students slept more than 8 hours.
7. No. 30% since young people nowadays do not have enough time to sleep.
8. Probably no, around 35% of students reported eight or more hours of sleep. Reason: college is stressful but people now tends to pay more attention on health

Student responses

1. Probably not, I don't think classes are in full swing and people are haveing to change there lsleep squedule to then get 8 hours now rather then letting it change

Student responses

1. 0.4. First week of school, less homework, more fun time.
2. 45%, College student dont really have enough sleeping time due to school work, internship, clubs and student jobs.
3. 65% because i believe it is more than half but only slightly since it is commong to get 8-10 hours of sleep and the ones dragging it down are all-nights and 3-4 hours
4. Given that this survey was likely to be answered the night before the first day of classes, I would say that a bit under 50% of students would get the recommended amount of sleep.
5. I estimate 55% of students slept at least 8 hours because its the begining of the semester.
6. i estimate that about 40% of the students got 8 or more hours of sleep because it's still the beginning of the semester and they likely don' tave much homework yet.
7. I think 40% of students sleep 8 or more hours. As the semester and school week picks up I believe more students will have seen their sleep time dwindle.
8. I think less than half go the respondent slpeoted eight or more4 hours last night because school just started and there's a lot going on

Student responses

1. I think that less than half of the students in this course got at least 8 hours of sleep. I think that about 40% got 8 hours of sleep because it is still early in the school year and school has not fully ramped up yet.
2. I think under half slept more than 8 hours, maybe 47%. Most uni students sleep less than 8 hrs.
3. I will guess about 60% of people got at least eight hours of sleep. My reasoning is that this was done before the first day of class, so people would not have been bogged down by work and would have less to do before sleeping. It may be lower if the stress of classes starting led to people having trouble falling asleep.
4. I would estimate roughly 40% because it is early on in the year, so people might be less busy but it is also recruiting season
5. I would estimate that slightly over half of students, 55%, get eight hours or more of sleep. That is because even though the average college student gets between 6-7 hours of sleep, this survey was created at the beginning of the semester before coursework has piled up.

Student responses

1. My guess is that 40% of students reported eight or more hours of sleep.
2. The distribution might be slightly skewed to the left, so I will say that it will be a little under 50%.
3. While on a normal school day I'd say no, I think it's a pretty close split as to how many people were partying on a weekend and then sleeping in. So I'd still say no, but probably 45% of people slept for at least 8 hours as we are still college students.
4. Yes, maybe about 67% percentage, since it's just the start of the semester, students will not have so much pressure about studying.

Student responses

1. I do not think at least half of students slept 8 hours or more. I think the majority slept less than 8 hours.
2. I do not think that at least of the respondents slept more than 8 hours. I think that 8 is a very high number for college students to have, especially considering many people in this class are upperclassmen and schedules have already started to pick up.
3. I do think that at least half the students sleep >8 hrs. This is the recommended sleep time for a normal lifestyle and I would estimate a 60%.
4. I don't think it was half, I started PS 1 and it looks more like 40%. But I would generally think it would be even lower (like a tenth) because of stereotypes about college students
5. I think 70% got at least 8 hours of sleep because it's syllabus week and we have less coursework right now.
6. I think it is yes but I personally only slept for 5 hours because of an online course. I guess 90% because it's the start and we are not busy

Student responses

1. I think more than half slept over 8 hours because the survey came out before school started so they had more time to sleep. It is also early in the year so more time to sleep.
2. I think that it is slightly skewed towards lower than 8 hours of sleep for college students, college is stressful and it is difficult to get 8 hours a night with that hectic schedule, so I think more than half sleep less than 8 hours
3. I think that more than half of the people slept at least 8 hours because it is syllabus week so people have less homework and will be able to go to bed earlier.
4. I think the average is about 7 hours of sleep because people may have gone to bed early in preparation for the first night of class, and even those who normally do not average 8 hours of sleep could have tried to get more sleep.
5. yes more than half of the class got 8 or more hours of sleep.

Student responses

1. 57%. Because i think at least a bunch of people haven't turned to study mode. So they didn't sleep enough.
2. 60%, because now it is just the beginning of the semester, work load is not so hard that can't get sleep for 8 hours
3. 60%; this was asked right before the start of semester, thus I'd guess people got more sleep than I was.
4. I estimate 55% reported eight or more hours because eight hours is the suggested amount of sleep every night
5. I think at least 50% of people sleep more than eight hours. I do not know all students' sleeping time, but I know many friends will sleep more than eight hours.
6. I think over half, probably around 60 percent or so, slept more than 8 hours. Because it is very early in the semester, people probably don't have much going on at the moment so they have more time to rest.
7. I think that at least half of the respondents got eight or more hours of sleep because most college students generally need around 7–9 hours of sleep each night.

Student responses

1. I would say that probably around 60% of people slept 8 hours due to the fact that the survey was on monday this week which was the first day of school so people where not very busy.
2. yes, maybe 70%. It's the beginning of the term so people don't have much work to do.

Student responses

1. I believe 50% or more of the students slept 8 or more hours in the survey because college classes hadn't started yet.
2. I believe that at least half of the class got eight or ore hours since it is only the first week of class and not much course work is due. Exams are not very close.
3. I predict that a little less than half slept 8 hours or more the night before. I estimate around a single digit percentage for the number of people who reported sleeping exactly (or estimated to) 8 hours the night before. The proportion is because of the environment. The percentage is due to an exact value.

Student responses

1. ,f college students have a lot of work to do
2. report more highgt
3. Yes?

What do you learn by summarizing `sleep\_value`?

99 anonymous responses

Mixed units / minutes vs hours

31

Many noted responses used different units (minutes vs hours), making summary numbers misleading and creating large max values.

Outliers and skew identified

21

Students observed extreme values (e.g., 510), outliers, and that the distribution is right-skewed or heavily skewed, affecting mean.

Descriptive statistics reported

15

Students said summarizing gives basic summary stats: min, max, mean, median, quartiles, IQR, range.

Median and its robustness

11

Comments focused on the median value (around 7.5) and that the median is less affected by outliers than the mean.

Response themes for question 2

Distribution / percentiles insights	10
Responses highlighting distributional interpretation: most people sleep 7–8 hours, quartiles, 75% have ≥ 7 or 8 hours, or general spread.	
Other responses	7
Responses the model could not place reliably.	
Unclear or miscellaneous comments	3
Responses that were vague, incomplete, or not directly about the summary content.	
Other responses	1

Student responses

1. I learned that the data is most likely in different units
2. I learned that the units are all not the same making the data be off
3. I learned that there was a difference in using mintues vs hours
4. I learned that we dont know what unit the hours are in
5. I've learned that unit of measurement could cause the data set to be unreliable.
6. It's meanningless because responses are in different units
7. mixed units identified
8. Not all of the sleep values are in the same unit, as the maximum is 510, which is probably minutes.
9. Peopl input their sleep duration in durrerent units

Student responses

1. People are using different units.
2. People reported their answer in different units. Some people did hours and others did minutes which is skewing results.
3. respondents mixed units when replying
4. Some people put it in minutes and some in hours, which is skewing the max and mean.
5. Some people recorded their sleep time in minutes which significantly raised the mean
6. Some people responded in hours while others responded in minutes.h
7. Some values are in minutes
8. That the data is not all reported in the same units
9. The data is in different units and needs to be changed to the same unit.

Student responses

1. The mean indicates that the units for all of the responses are not standardized.
2. The units of the data might be mixed up, like mixing mins and hours
3. there are different units submitted
4. There are possible outliers. If people used different units to describe the amount of sleep they got. The max, which is 510, was likely minutes.
5. There are responses measured in minutes that skew the mean and max values
6. There is a huge range of answers (min = 2 and max = 510). Probably because of different units (minutes vs hours)
7. there's a different set of units for responses
8. unit of measure is not the same in the data set.
9. we didn't learn anything because the values in the data have different units that make the summary inaccurate

Student responses

1. We have mixed units of minutes and hours.
2. We have the same problem as with height where people used different units so the summary command isn't giving us any useful information.
3. We learned that it is likely that not all measurements in the data set are measured in hours because the mean is so high
4. While it usually gives a range, minimum, maximum, 1st and 3rd quartile, median and mean, here it includes minute values which we have not yet accounted for, so many of these values are off.

Student responses

1. data is not cleaned wit the max being 510
2. Distribution is heavily skewed to the right
3. I learned that the data is likely in two different units (minutes and hours) due to the vast range of output.
4. I learned the range of responses was 2 to 510
5. It is a right skew curve from the value.
6. Minutes in the dataset are an outlier and need to be accounted for. median is below 8 which is an okay repesatation
7. that people are mainly getting 7 to 8 hours of sleep. but the max value seems to be skewing the data
8. The 50th percentile of the distribution is 7.50
9. the data is heavily skewed to the left

Student responses

1. The data is significant right skewed, but consider the unit is different can be explained.
2. The mean > median, so the data is right skewed
3. the mean is extremely high which means there is unusual data in the data set
4. The mean is much higher compared to the median of sleep values
5. The median is much lower than mean, so there may be some outliers
6. The skewed right because mean is higher than median.
7. There is an impossible sleep value response, an outlier of 510 hours, skewing the distribution to the right
8. We find that there is likely an outlier, due to the mean being drastically different from the median, and the fact that the max is 510.
9. We learn that the mean and median are off there is a weird spread in the data

Student responses

1. We learn that this data has the same problem as last time with some units described in minutes which skews the data
2. We learned that not everyone measured their sleep in hours, as the mean of 26.28 and the max of 510 show.
3. YOu learn that there are extreme outliers on the top end of the disribution.

Student responses

1. By summarizing data, you learn the IQR values of the sleep data, like the minimum, maximum, quartiles, etc.
2. general distribution of the data (min, max, median)
3. i learned that it gave useful information like min and median
4. i learned what the median of the whole sleep data set was
5. IQR, range, median
6. It gives you a lot of important info about the data spread
7. statistical dist. of values from the given sleep data
8. The main stats of the data
9. the spread of the data and important values

Student responses

1. We can learn the min median mean and max but since the units are unknown the numbers won't make sense.
2. We learned key statistics based on the raw data before preprocessing, such as mean and median.
3. You find the minimum, max, and other things that are useful to know about the range of data.
4. You learn how the data is distributed through Q1, Q3, and median, as well as the range.
5. You learn the min, 1st Qu., median, mean, 3rd Quarter and Max
6. You learn what the min, median, mean and other variables are

Student responses

1. Merely from the value and without considering the unit, median is 7.5 hours. But we still have to consider units (min) / hrs etc.
2. People used different units to report their sleep and the median is similar to minimum recommended amount of sleep for adults.
3. Sleep data was recorded mostly in hours, but some in minutes, as shown by the high max value. The median is resistant to these outliers, and is probably mostly correct in showing that the average student got 7.5 hours of sleep.
4. The median is lower than our null hypothesis of college students getting 8 hours of sleep a night
5. The median sleep number is less than 8
6. the median will not be affected by those outliers
7. We learn that the median amount of sleep per night was 7.5 hours.
8. we learn that the median amount of sleep was 7.5 hours, but that the data may have been skewed by using different units, such as minutes
9. We learned from purely the numerical data, that the average person is sleeping 7.5 hours. The median is resistant to outliers.

Student responses

1. We learned that the minimum amount of time that someone who answered the survey slept for was 2 hours.
2. We understand that due to a lack of standardised units, mean does not work well. The most likely number is 7.5 hours.

Student responses

1. 75% of people got 7 or more hours of sleep.
2. I guess we learn how to summarize the vectors of weird spread data.
3. Most students report their hours of sleep in hours, considering how the 3rd quartile says "8", at least 75% of the students who completed the survey reported their sleep time in hours
4. That most people are getting between 7 and 8 hours
5. that roughly 25% of people slept 8 hours or more
6. We can learn that we can see the range.
7. We learn how the data might be distributed. We also see we need to standardize data
8. We learn that the majority of students used hours to answer, but some used minutes, based on the fact that Q1, the median, and Q3 are all reasonable hour answers, but the maximum is 510.
9. You can see that most of the data is around 2 - 8, with some much higher values (which were submitted in minutes)

Student responses

1. You learn the general idea for the sleep amount distribution

Student responses

1. I think the data is skewed because people used different units of measurements when recording their sleep.
2. That less than 50% got more than eight hours. But that the data is both in minutes and hours due to the high mean.
3. the max is 510, so there are mixed units (someone probably didn't sleep 510 hours)
4. The mean or median of the variable
5. The median of sleep hours
6. there is a very extreme value
7. Use select can create a small sheet

Student responses

1. I think the data will be left skewed because
2. the column row and values of sleep value
3. There might e something rong?

Combine ``mutate`` and ``filter`` to determine how many hours survey respondents ``R016`` and ``R031`` slept.

82 anonymous responses

Filter usage examples / code snippets

26

Short code-focused attempts showing `filter` or `filter` with selection of `response_id`, sometimes syntactically incorrect but targeting filtering by id.

Correct approach: mutate then filter / pipe

22

Answers stating to convert units or create a new column with `mutate` (e.g., `minutes` → `hours` or `got_eight_hours`) then `filter` for the two `response_ids`, often recommending the pipe.

Correct approach: filter for IDs then mutate

15

Students suggesting to first `filter` for the specific `response_ids` (`R016`, `R031`) and then `inspect` or `mutate` to get hours.

Direct numeric answers

11

Students who gave specific sleep-hour values for `R016` and `R031` (numeric results).

Response themes for question 3

Misunderstanding or incorrect operations	4
Responses that misunderstand combine/filter usage (e.g., summing two rows, filtering ranges incorrectly, or odd syntax).	
Other responses	4

Student responses

1. 13? Trying to figure out if I combined it correctly
2. 463, 510
3. `df <- mutate(filter(df, got_eight_hours))`
4. `eight_hour_sleep <- filter(df, got_eight_hours)` `df <- mutate(df, eight_hour_sleep = sleep_hours >= 8)`
5. filter the table twice.
6. filter by response_id
7. filter by the two rows that have the response ID r016 and r031
8. filter by response_id only
9. filter for the names using mutate and filter

Student responses

1. Filter out rows R016 and row R031 and then add the two hours slept together to get the total.
2. filter response_id
3. filter the rows that only have the listed response ids and to do that we need the conditional statements to do that
4. Filter the table twice and find the rows of data (condition) where the respondent ID is R016 and R031. filter(respondent_id = R016 & respondent_id = R031)
5. filter(df, R016) filter(df, R031)
6. filter(df, response_id == "R016")
7. filter(df, response_id=="R016")
8. I honestly dont know
9. Isolate the two rows using mutate and filter the table

Student responses

1. `mutate(df, got_eight_hours, filter(df,`
2. `mutate(eight_hours_sleep, sleep_value) filter(eight_hours_sleep, in (R016,R031))`
3. Mutt so it contains to row where response id is 16 and 3ate
4. `respondent <- combine(id_R016, id_Ro31); df <- mutate(df,`
5. `sixteen_and_thirty_one <- filter(df, response_id == "R016", respondent_id == "R031") sixteen_and_thirty_one`
6. `sixteen_and_thirtyone <- filter(df, response_id == R016 || response_id == R031)`
7. Use pipe
8. Use pipe to filter

Student responses

1. `df <- mutate( sleep_data, sleep_hours = if_else( sleep_unit == "minutes", sleep_value / 60, sleep_value ) ) filter(respondent_id = R016 & responent_id = R031)`
2. `df <- mutate(df, got_eight_hours = sleep_hours >= 8) eight_hour_sleep <- filter(df, response_id == "R016" | response_id == "R031")`
3. `df <- mutate(df, respondents = response_id == "R016" | response_id == "R031")`
4. `df_spec <- mutate(df, spec= (response_id == "R016" | response_id == "R031")) df_spec spec_hrs <- filter(df_spec, spec) spec_hrs`
5. Filter out all the other survey respondents while mutating the remaining two to get sleep time
6. Filter the column response_id to get just the entries from 16 and 31. Mutate to create a new column combining the number of hours slept and the word hours?
7. filter to include the two rows then mutate to isolate the sleep hours value
8. filter to keep R016 and R031, then mutate to return the sleep values
9. filter we use to filter by column of the dataframe to get the rows. mutate to convert minutes to hours

Student responses

1. filter() and then mutate()
2. I can use a pipe. I want to mutate and then use filter. `filter(df, response_id == "R016")`
3. isolate two rows with r016 and r031 by filter
4. make a new column like `sleep_hours = minutes / 60`, then select the rows where the respondent ID is R016 or R031. The result should show $463/60 = 7.72$ hours for R016 and $510/60 = 8.5$ hours for R031. If you want, I can also show the exact dplyr code.
5. mutate df to get mask, then filter on mask
6. mutate first and then filter
7. Mutate the data through pipe
8. Mutate, then filter twice
9. To do this you would first need to get everything into the basic unit of hours and then from there you can filter it so that you are only seeing the response_id that is for R016 and R031

Student responses

1. Use filter to specify response id and then mutate into new column and print
2. Use mutate to do the same in the past few slides, then use filter to select response_id R016 and R031.
3. use the pipe to mutate first and then filter
4. You need to first isolate those two respondents using response id, which will be don using filter, then use mutate to get the number of hours they slept

Student responses

1. 8 hours
2. Filter out rows 16-31
3. Filter out the two respondents and then find how many hours each one slept
4. filter the data frame to show rows of data that correspond to R016, and then do the same thing for entry R031
5. filter the data to show only R016 to R031 and then mutate the data to show the hours of sleep for those
6. Filter the table for response_id == "R016" or "R031"
7. filter(df, response_id == "R016")
8. I used filter(df, response_id == "R016") and R031 to determine it.
9. lows = filter(df, response_id == "R016" | response_id == "R031") select(lows, response_id, sleep_hours)

Student responses

1. `ro16_ro31 <- filter(df, response_id=="R016" or response_id=='R031')`
2. `sixteen <- filter(df, response_id == "R016")` `thirtyone <- filter(df, response_id == "R031")`
3. `surR016` slept 7.72 hours, and `R031` slept 8.5 hours
4. use filter
5. Use the filter command on the response id
6. You can filter with a `or` function as the clause to identify the two ids, or run filter twice.

Student responses

1. 7.5 8
2. 7.71 and 8.5
3. 7.72 8.5
4. 7.72 h and 8.5 h
5. 7.72 hours for R016 and 8.5 hours for R031
6. 7.72, 8.5
7. 8.5,7.5
8. R016 slept 6 hours R031 slept 7
9. R016 slept 7.717 and R031 slept 8.5

Student responses

1. R016 slept 7.72 hrs and R031 slept 8.5 hrs
2. R016: 7.72, R031: 8.5

Student responses

1. 8 hours
2. Both greater than 8
3. `df <- mutate(df, sleepdata, ifelse(responsive data is r016 and 4031) then I can filter on those`
4. You would need to filter and crop them out for a articular id and it will give aroun 7 hours

Student responses

1. filter by response id
2. find the rows for respondents r016 and r031 and find the value for hours slept
3. I dont know
4. Not sure the exact syntax but I have to filter for the respondents

Starting from the data frame we just created, explain what additional steps you need to take to answer the question, "Did at least half of the respondents sleep eight or more hours?"

103 anonymous responses

Compute proportion and compare to 50%

58

Divide the count of respondents with eight+ hours by the total sample size (e.g., 44/98) to get a percentage or proportion and compare it to 0.5.

Filter or count respondents meeting condition

29

Describe filtering the data frame for `sleep_hours >= 8` or counting TRUE values in `got_eight_hours` to get the number of respondents who slept eight or more hours.

Other responses

11

Responses the model could not place reliably.

Ensure consistent units or data verification

3

Mention verifying or converting units to hours and checking data quality before computing counts or proportions.

Response themes for question 4

Other responses

2

Student responses

1. .4489796
2. 44
3. 44 people out of 98 slept 8 hours or more, which is 45%. I did this by filtering the data frame to only include rows where the respondednt slept more than 8 hours.
4. 44.9%
5. 44/98
6. 44/98 got eight hours of sleep. I found this anwer by summarizing the dataset.
7. All of the data must be in the same unit of time
8. Calculate the percentage of sleep values that are greater than or equal to 8 amongst the entire dataset.
9. Calculate the proportions of respondents who satisfy the needs of the research question

Student responses

1. Compare half of the respondents to the number of rows in eight_hour_sleep.
2. compute proportions by dividing total
3. compute the proportion of people who slept eight or more hours
4. count people who slept more than 8 hours, divide by total
5. count respondents with more than 8 hours slept and divide that by verall sample
6. Count the number of respondents who slept at least eight hours by computing and dividing by total responses
7. Count the number of TRUE responses for eight hour sleep column and then divide that by the toal to get the proportion
8. `count(eight_hour_sleep) / count(df)`
9. cupute properties buy divinh total

Student responses

1. Determine the size of the vector that contains got_eight_hours of sleep with the true condition which is 44/98 which is 45% percent.
2. Divide by original dimension
3. divide by the total amount of respondents and see if its greater than 50%
4. Divide the number of people we found responded that they got 8+ hours by total number of responses
5. divide the number of responses that got at least 8 hours of sleep by the total number of responses, so 44/98 which is 45%
6. Divide the number of true responses by the total number of survey respondents, or 44/98. This gives us a value of 0.449, or roughly 45%, which is not a majority of the class.
7. divided by the total number
8. filter and count the number of respondents who slept eight or more hours
9. find the total number of people where got_eight_hours is true, then divide it by total number of respondents

Student responses

1. First, I would filter the data to respondents who slept 8 or more hours, count them, and then divide that number by the total number of respondents to find the percentage.
2. Get the length of the vector of people who got eight hours of sleep or more, and then divide that number by the total length of the total vector and compare it to 50%
3. Get the length of the vector of the peopel who got 8 hours of sleep, more than/equal to 8 horus of sleep*.then divide by all respondents
4. I don't know how to do it.
5. `length(filter(df, sleep_hours >= 8)$sleep_value) / length(df$sleep_value) = 44.9%`
6. `n = count(filter(df, df$sleep_hours > 8)) n`
7. Need a step to calculate the total number of respondants and the number of responders with eight or more hours ofsleeps so that we can calculae the percentage.
8. No
9. No, about 44%.89 got 8 hours of sleep

Student responses

1. No, less than half of the class got 8 hours of sleep. Only .449 of the class did
2. `nrow(df |> filter(got_eight_hours == "TRUE")) / nrow(sleep_data)`
3. `nrow(eight_hour_sleep) / nrow(df)` gives 0.449 which means less than half the responses slept over 8 hrs
4. `nrow(eight_hour_sleep) / nrow(df)`
5. `nrow(filter(df, got_eight_hours)) / nrow(df) = 0.448`, so no, less than 50%
6. Set the number of respondents who slept eight or more hours over the number of respondents to see the fraction
7. Since all our responses are in the same units now, we can compare the proportion of respondents who slept more than eight hours with 50percent and see if it is higher
8. Since there are 98 responses, then we could sum the number of people who have a "TRUE" value for `sleep_over_eight` by filtering, and, if it is more than 49, then over half would have slept over eight hours. However, the answer is 44, which means less than half got at least eight hours of sleep.
9. tally up the number of TRUE values in the col that indicidated whether or not the response got more than 8 hours of sleep, then divide by 98

Student responses

1. There are 98 respondents, we could mutate a boolean to true when they slept over 8 hours and sum that
2. There were 44 people that did sleep over so we divide that by 98 and we will receive our total which is less than 50
3. Use filter to do True or False to get precentage, and compares that with 50%
4. Use if else and some nrow/2 to make sure that's a half. And then compare with `sum(sleep_hours > 8)`
5. using `summary(df$sleep_hours)`, found out that mean is 7.344 so the average is not more than 8 hours
6. we can filter out responses tat got more than 8 hours and then divide by 98 so you would get 44/98 got more than 8 hours of sleep
7. we know 44 respondents slept 8+ hrs so divide it by the total number of respondents
8. We need to calculate the proportion of respondents sleeping eight or more hours, which first requires having the total number of respondents
9. We need to filter for the respondents whose "got_eight_hours" value is TRUE, which is 44 people.

Student responses

1. we need to find the proportion >8 for sleep_hours
2. We need to know how many total respondents there were. Then we need to determine how many people slept eight or more hours and put this number over the total to get the percentage
3. we would need to check the percentiles of people who slept more than 8 hours and those who did not after converting all units to hours
4. You would take the count of the kids who slept 8 or more hours and then divide that by the total number of responses

Student responses

1. 47
2. count the amount of true in the mutated data and then compute the proportion
3. Create a boolean expression to see if 8+ hours is true for the data set and then only print the ones who slept more than 8 hours, we can then divide this by the total number of respondents
4. create a counter variable and add one everytime the got eight hours column is true
5. filter
6. Filter df by the 'got_eight_hours' column being true, then compare the number of rows in the resulting table with the total number of respondents
7. filter for equal to or more than 8 hours
8. filter got_eight_hours to true and then count how many rows are in that column or getting the size/length of that filtered column
9. filter to sleep hours ≥ 8 and see how many rows pull up and divide by 98

Student responses

1. `filter(df,sleep_value>8 )`
2. Find the actual number of people that slept 8 or more hours? So maybe use the length function to find the length of the modified dataframe?
3. For all people marked with True, for more than 8 hrs count that and then check out of total people, greater than 50%
4. from the dataframe where we filtered for people who got 8 hours or more, count how many rows and compare to the # of rows from our initial data frame
5. Iterate through all the `got_eight_hours` and add 1 each time the answer is true. You can divide the number of true values by the total number of respondents.
6. Just count how many `response_id` have a True value in `eight_hours`
7. make a table off of the data frame got eight hours
8. Use `nrow` on the filtered, `eight_hour_sleep` dataframe.
9. we can use `dim(eight_hour_sleep)` to find the number of people who got more than 8 hrs, and then divide it by the number of responses

Student responses

1. We might filter the data and make sleep ≥ 8
2. we might want to filter the data to only 8 or more hours and then count the number of respondents this applies to
3. we need to count the respondents with sleep_hours ≥ 8 , compute the percentage and compare it to 50%
4. we need to create a variable that can count got_eight_hours == "TRUE"
5. We need to filter out all the responses that didn't get 8 or more hours of sleep, and count them. Then we can compare that number to the total number of respondents.
6. we need to run more tests by computing how many respondents reported at least 8 hours
7. We would have to count the number fitting the condition of the hours
8. would need to filter df by True for slept_eight_hours column. Then would would to count the rows in this filter
9. you can create a count and add 1 to it each time there is an entry in eight_hours_sleep

Student responses

1. You need filter the number of people with more than 8 hours of sleep in the data frame then count the number of rows in that df
2. You would need to filter the df on true for go_eight_hours then divide be sample size

Student responses

1. 43/98 responses reported more than 8 hours. You can filter anything that has sleep hours more than or equal to 8 and see how many responses are left
2. Count the number of rows (44) in eight_hour_sleep
3. Count the number of TRUE responses for eight_hour_sleep column, divide that by the total to get the proportion, compare to 1/2, 44.89%
4. Loop through sleep_data and count the number of response with TRUE in got_eight_hours.
5. Maybe try to summarize all the numbers of people who sleep over 8 hours, then add them up, finally check if the number is over 50% or not?
6. No, 42/98 students slept 8 or more hours.
7. only around 16.5%
8. Use an if_else command to separate respondents who sleep more than 8 hours or less.
9. You can use the filter command to filter out how many columns of people slept over eight hours and then divide that number by the total

Student responses

1. you could use boolean logic and some sort of for loop and count the number of responses where sleep_hours is greater than 8
2. You need to say d or count the number of true responses in the table after asking for the number of responses with a value

Student responses

1. Other steps is to verify our data and make sure all the units of measurements are the same
2. You would have to filter the data to give responses in just hours, then get the mean of this data and divide it by the number of rows. If the output is 8 or above, then the answer would be yes.
3. You would need to convert all of the answers to the same unit (hours) and then divide by the total # of responses (98) to see if a majority of the observational units slept more/less than 8 hours

What percentage of survey respondents slept at least nine hours? What percentage slept fewer than six hours?

0 anonymous responses

No responses were submitted.

Among the 98 survey responses: 1. Which favorite food do you predict was selected most often? Why? 2. Which continent of birth do you predict was reported most often? Why?

101 anonymous responses

Predict pizza and North America

59

Students predict pizza is most common and North America is most common birthplace, often citing campus location, American student majority, or pizza's popularity.

Predict sushi and North America

23

Students predict sushi most common for food and North America most common birthplace, citing sushi's popularity and local student demographics.

Pizza predicted; continent unclear

9

Students favor pizza but give little or no reasoning about continent or leave continent unspecified.

Response themes for question 6

Other foods predicted with North America

6

Students predict a different food (pasta, burger, chips, pho, garlic, rice, hamburgers) but still expect North America most common birthplace due to university location or demographics.

Other responses

4

Responses the model could not place reliably.

Student responses

1. 1. I predict more than half will select pizza. I predict more than half will report North America
2. 1. I predict pizza was the most popular 2. I think North America was the most popular place of birth
3. 1. I predict that Pizza is the most common food selected because it is the most universally loved. 2. I predict North America was the most common continent selected as our class is based in North America.
4. 1. I think most people selected Pizza 2. I think most people were born in North America
5. 1. I think pizza was the most chosen answer because it is a widely liked food. 2. North America would be the most common continent because the university hails from Michigan
6. 1. I would say Pizza as it is pretty universal and most people from america in this room. 2. I would say america north
7. 1. pizza 2. north america
8. 1. Pizza 2. North America
9. 1. Pizza 2. North America because instate and out of state is way more than international population

Student responses

1. 1. Pizza 2. North America because Michigan is a school inside North America.
2. 1. pizza 2. north america because we are in the us
3. 1. Pizza 2. North America, the uni is in the u.s so whole expect most to be domestic Students most likely born in NA, pizza is common
4. 1. Pizza 2. North America, the US is in North America
5. 1. Pizza 2. USA
6. 1. Pizza as people like it. 2. North America as this school is in North America.
7. 1. Pizza because it is a really popular american food. 2. North America because we are in school in north america
8. 1. Pizza because it is readily abundant and easy to obtain 2. North America because this university is located in North America
9. 1. pizza because it seems like it would suit the tastes of 20 year olds 2. north america because most people attending uMich are from the US

Student responses

1. 1. Pizza because most available food 2. North America, location of UMICH
2. 1. Pizza was most likely the most common because it is very popular on campus, and there are a lot of pizza restaurants. 2. North America because most people who attend U-M are from this continent
3. 1. Pizza, because it is a very typical staple American (Italian) food. 2. I predict the most common continent of birth is North America because UofM is in North America.
4. 1. Pizza, its extremely common worldwide 2. North America, since thats the continent we are in
5. 1. pizza; 2. north america
6. 1. pizza. 2. USA because they are both americian style.
7. 1.pizza, 2.North America
8. 1) Pizza because that is a popular food. 2) North america because i think the most people were born there
9. I do not remember the options for food. I think north America was most likely because we are at the university of michigan.

Student responses

1. I predict pizza was chosen the most often, because it is a very common favorite food. I think the Americas is the most popular birth continent, as this school is located in the U.S.
2. I predict pizza was selected most often because it is a common food with minimal ingredients and I predict North America was most reported
3. I predict that pizza was the number one prediction for food because it's extremely popular, and i predict that north america was the most popular continent of birth because it is where michigan is, where we all go to school
4. I predict that pizza will be the most common because I think that is the most standard. I also think North America will be most popular because we are in North America
5. I thin Pizza was probably selected most often because it is just a generally popular food and then North america is probably the most common birth continent because UM is located in North America
6. I think pizza and North America were most common
7. I think pizza bc is it is the most common food, US because lots of people are from here and they like pizza

Student responses

1. I think Pizza because it is commonly ate around the US and North America because we live there.
2. I think pizza was selected most just because it seems the most universally liked, especially among college students because it's cheap and accessible. I think North America is the most common location of birth because we're located at an American university and most people here were born in America
3. I think pizza was selected most often because it is a largely popular food in the area and I think most people are from north america because that's where we are located.
4. i think Pizza, because that is the most basic and common answer that is popular among average people North america was definitely the most common birth continent because most people here would be from the US
5. I think Pizza, it's pretty popular. I think N would be most common because we're in USA.
6. I think that pizza was the most selected answer for food and North America for continent
7. I think the favorite food that was selected most often was pizza and the most common continent was North America.

Student responses

1. I would say the pizza was the most selected one because its the most common food in this continent. Going from that, I predict that the north america was the continent of birth chosen the most
2. Pizza and north america
3. pizza and north America because pizza is very popular and we go to mchigan;
4. Pizza and North America, regional
5. Pizza since it is very popular, NOorth america since we arein America
6. Pizza was the favorite because it is partially all i Remember and also it sees to be the least contested. I think the America will be the most common one as his is not a school with >50% immigrants or anywhere close.
7. pizza, anecdotal most common answer 2. North america because that it where university of michigan is located
8. Pizza, because a majority of the class was likely born in the united states considering half of the university is from in-state.
9. Pizza, because it seems to be a popular food and North America since most students that go to college in the U.S. wer born in the U.S.

Student responses

1. Pizza, because most people were likely born in North America/the U.S., and it is an American staple. I expect NA to be the most reported continent of birth.
2. pizza, north america
3. pizza, North america
4. Pizza, North America UMich is a US institution so most of the students would be born in North America, where Pizza is a popular food
5. Pizza; super common food North America; we're in north america
6. Pizza? At top of the selection. Continent maybe the North America
7. Pizza. Because pizza is popular. North America. Because U of M is a school in North America
8. pizza. This is based on the assumption that most people are born locally, and most americans would choose pizza.
9. Sushi because it seems to be very popular. North America because most people are from the US at UM

Student responses

1. 1. I predict sushi was selected most often because most of my friends like sushi the most. 2. I predict North America was most common because Michigan is in North America.
2. 1. I think sushi was picked because I think its very popular among states, countries etc. It can be changed for whatever culture/country. 2. North America because we are in North America, so I assume most students are from here with the exception of some.
3. 1. My prediction is that the favorite food selected most often was sushi, because I know that's a commonly liked food. 2. I predict North America to be the continent in which majority were born because we are in NA.
4. 1. sushi 2. america
5. 1. sushi 2. usa
6. 1. Sushi because its very popular. 2. North America because we are in North America
7. 1. sushi because that's my favorite 2. North America because that's where we attend University
8. 1. Sushi, as its popular but not something that people eat all the time. 2. North America, because the majority of students at Umich are from the US.
9. 1. sushi? because that's what I responded 2. North America that's where aa is at

Student responses

1. sushi just generally good at everything idk 2. probably pizza in america
2. 1) I think the most common favorite food is sushi, cause it's basic 2) America because the school is in America
3. 1) Sushi, you can always have fresh sushi and u can have it for any meal. it also is semi healthy 2) North America, we are in north america
4. I predict sushi because I think everyone loves sushi based on my life's observations. I think the continent of birth most often was North America
5. I think favorite food is sushi due to. America is reported most as we in the US.
6. I think Sushi was selected most often because I think I saw it as the answer quite a few times. I think North America is the most popular continent of birth since we're here right now.
7. I think sushi was selected most often because just thinking of the demographic of people I know from UMich, it's definitely the most common. North America was definitely the most common continent of birth, since there aren't that many international students.

Student responses

1. I think that pizza or sushi was selected most often, because they are common, customizable foods that can be catered to many tastes. I think America is the most selected continent of birth, because we are in America.
2. sushi because I glimpsed at the data, North America because that's where the university that took the survey is
3. Sushi because people like expensive nice things, and North America because we are in the USA
4. Sushi or Burrito, as they are very popular. North America, because USA.
5. Sushi, North America
6. Sushi, people like sushi, I think North America is most common because we are at a north american university
7. sushi, random guess america, the uni is in america

Student responses

1. 1. burger 2. rice, because Asian eat rice, because most students are white
2. 1. Chips. Lots of Americans love it. 2. North America. This is an American university so without doubt dominated by native
3. 1. i admittedly don't remember the favorite food options 2. North America, considering that it is where we are located right now
4. 1. pizza 2.
5. Garlic, because I love garlic North America, since it's in the university
6. i say burger because it is standard. north america
7. I think that ramen may have been the favorite food for people given it is college students. I think North America was most common
8. pizza, because it is the symbol food of US
9. pizza, since that's a american party food

Student responses

1. 1. Pasta and Italian food because it is very common 2. North Amwrica
2. hamburger. north america
3. hamburgers, because they are delicious and unhealthy. Asia because two countries with the biggest population are in Asia
4. I think italian because its very common and north america because umich is in the US so most students were likely born here
5. Pho, because its yummy. and NA because UMich is in NA
6. rice cuz every eat them? I guess Asia have the most birth since russia, india, china.

Student responses

1. i think most people would be born in north america
1. Pizza, its the classic food that people eat and is cheap 2. North America, there are more US born students at Mich than International
1. sushi 2. america
- The united states was the most common and I think most people put pizza or sushi because there are a lot of great pizza and sushi places in Ann Arbor

What percentage of survey respondents were born in Asia? North America? Elsewhere? (Hint: use the `table()` function and divide by the total number of respondents.)

97 anonymous responses

Precise correct percentages

49

Students stating the calculated percentages matching 18/98 (~18.4%) for Asia and 80/98 (~81.6%) for North America, often noting 0% elsewhere.

R code fragments or formulas

20

Students providing code-like expressions or calculations (sums, ratios, or function fragments) rather than final percentages.

Approximate or differing numeric guesses

15

Students offering rough estimates or alternate percentage splits that differ from the 18/80 breakdown (various guesses for Asia, North America, and elsewhere).

Method: use `table()` and divide

11

Students describing the correct procedure using `table()` or similar R commands to count and divide by the total (n or 98).

Response themes for question 7

Other responses

2

Student responses

1. 0.18, 0.81, 0.00 elsewhere
2. 0.183673469
3. 18 were born in Asia and 80 are born in North America.
4. 18.3 in Asia and 81.6 in North America,percentage wise
5. 18.36% were born in Asia and 81.64% were born in North Ameria 0% were born elsewhere
6. 18.37% from Asia 81.63% of North America no other continents are represented in the survey
7. 18.4 and 81 percent
8. 18.4 percent in asia, and 81.6 percent in North America
9. 18.4%

Student responses

1. 18.4% Asia, 81.6% North America, 0% elsewhere
2. 18.4% for Asia, 81.6% for North America
3. 18.4% in Asia, 81.6% in North America, and 0 elsewhere
4. 18.4% were born in Asia and 81.6% were born in North America.
5. 18/98
6. 18/98 and 80/98.
7. 18/98 from Asia. 80/98 were from North America
8. 18/98 in Asia, 80/98 in North America, 0% everywhere else
9. 18/98 of survey respondents were born in Asia, 80/98 of survey respondents were born in North America.

Student responses

1. 18/98 were Asia and 80/98 were North America. This is ~18% and ~82% respectively
2. 18/98 were born in Asia and 80/98 were born in North America
3. 18/98 were born in Asia and 80/98 were born in North America. No other continents
4. 18/98 were born in Asia, while the other 80 were born in North America. No other continent was reported.
5. 18/98 were from asia so 18.4 percent
6. 18/98, 80/98 born North America, else where 0.
7. 18/98, 80/98, Asia and North America respectively
8. 18/98=0.186 (Asia
9. 18% asia and 82 for North America

Student responses

1. 18% of respondents were born in Asia, while all the rest are from North America.
2. 18% of respondents were born in Asia and 82% were born in North America with no respondents being born elsewhere.
3. 18% of respondents were born in Asia, and 82% were born in North America
4. 18% were born in Asia; 81% were born in American (roughly)
5. 80 out of 98 (81.6%) were born in North America, and 18 out of 98 (18.4%) were born in Asia
6. 80 people born North America and 18 in Asia so those would correspond to 80/98 and 18/98
7. 80/98 were born in North America and 18/98 were born in Asia
8. 80/98 were born in North America, and 18/98 were born in Asia.
9. 81.63% from North America, remainder from Asia. Nowhere else.

Student responses

1. 82 % in North Americna and 18% in asia
2. 82% North America & 18% Asia
3. asia - $18/98 =$ north america - $80/98 = 81.6\%$ 18.4% for asia
4. Asia is 18% North America 81%
5. Asia: 18.5% north america: 81.5%
6. Asia: $18/98 = 18.37\%$ North Ameica: $80/98 = 81.63\%$
7. asia: 18/98 north america: 80/98
8. asia: 18/98 north america: 80/98
9. Asia: 18/98 or around 18 percent North America: 80/98 or 81 percent Elsewhere: 0

Student responses

1. i say 18.4%
2. North America = 80/98 Asia = 18/98
3. North America 81.6%, Asia 18.4%
4. There is 18.4% of respondents in Asia, 81.6% in North America, and 0% elsewhere.

Student responses

1. .816 US and .183 Asia
2. 10%
3. 18.4%/81.6%
4. 19%
5. 19% in Asia 81% in North America
6. 25%, 60%
7. 25%of Asia, 40%of US,
8. 6
9. 60% North America, 15% Asia, and 25% other

Student responses

1. 80% north america
2. 80%north america and 20% asia
3. Asia - 20% North America - 80%
4. Asia = 7%, North America = 90% elsewhere = 3%.
5. asia 40% north america40%
6. asia: 0.1836735 0.9183673
7. Asia: 4/10, North America: 5/10, Elsewhere:1/10
8. filter(tabel(respondent) born in asia/ total
9. I believe that 90% of the people in this class are from North America. I think that 8% are from asia because this school is popular there and datasci is dominated by asian/white men at umich. Last 2% is everybody else.

Student responses

1. My Code isbr working
2. roughly 80% in North America and 11% in Asia and 9% somewhere else

Student responses

1. 10%,70%,20%
2. 15%
3. 18% 80% 2%
4. 18% are born in asia, 82% are born in america, with no one being born elsewhere.
5. 18% in Asia; 82% in North America
6. 18%, 80%
7. 19 percent of respondents were born in Asia and 81 percent were born in North America
8. 20 percent Asia and 80 percent North America
9. 20%

Student responses

1. 80% north america 15% asia 5% elsewhere
2. about 80% of people from North America, and the other 20% from Asia
3. Around 80/98 were born in North America. Around 17 percent in Asia.
4. Asia 30%, 50% north america, 20% elsewhere
5. Asia: 30% North America: 10%
6. I would guess 20% were born in Asia, 80% in North America,

Student responses

1. Divide the people from asia by the total
2. `filter(survey, birth_continent == 'Asia')`
3. `filter(survey, birth_continent == "Asia")`
4. `sum(survey$birth_continent == "Asia") / 98`
5. `table(survey$birth_continent == "Asia")`
6. `table(survey$birth_continent)` count then divide
7. `table(survey$birthplace)`, count and divide by hand or by `nrow(survey)`
8. `table(survey$born_continent) / total_respondents`
9. `table(survey$continent_born)/98`

Student responses

1. `table(survey$continent)` then filter out responses for a continent and divide by total
2. Use table and then divide by 98. so `table(survey$colname) / 98`

One graph shows food choices for the whole class. The other shows continents of birth for the whole class. Using both these graphs, can we tell whether food choices differed between respondents born in Asia and respondents born in North America? Why or why not?

99 anonymous responses

Cannot tell from separate marginal plots

84

Students state that because each graph shows only overall distributions (marginals) and not which respondents belong to which category, the joint distribution or correlation between birth continent and food choice cannot be determined from the two graphs alone.

Need joint/filtered analysis

6

Students say you could answer the question if you filtered or re-plotted the data to show food choice by continent (e.g., create a joint distribution, separate bar charts, or run a test).

Other responses

4

Responses the model could not place reliably.

Response themes for question 8

Brief affirmative that differences exist

3

Students assert that differences exist between respondents from different regions or that region predicts food preference, without noting missing joint data.

Other responses

2

Student responses

1. Based on this graph probably not. There isn't enough data to answer this question
2. cannot tell from separate graphs because data are not related to each other
3. From looking at the graphs you can't tell any relationship between the two variables. In order to show a relationship we would need to do further analysis to see if there is a relationship
4. No because because we don't have enough information
5. No because by these tables we can't see what people from North America or Asia have picked for food choices
6. No because there is no exact way where we can visualize where both variables are connected
7. No because there is no information given
8. No because they are separate graphs.
9. No because they are two separate questions

Student responses

1. no because we can't identify which respondents said what from the graphs so we dont know if there is any correlation between birthplace and favorite food.
2. No because we can't see birth continent and food choices at the individual level. Can't see how both variables relate to each other since we're seeing it marginally
3. No because we do not know how they match together
4. No because we would have to get information showing how many people said a certain food that are from one or the other continent.
5. No the marginal distribution of the food preferences does not directly correlate to respondents born in North America and Asia
6. No we can not because they are independent variables
7. No we can not because those two variables are grouped together and we dont know how they relate
8. No we can not because those two variables are not grouped together to see how they relate
9. No we can't because there is no way to tell what data is within what continent

Student responses

1. no we can't tell if theres an effect on food choice by birth continent because the data doesn't account for a respondents birth continent when displaying food choice.
2. no we cant cuz we dont know who submit for the foods and continents of birth answers.
3. no without seperation we cannot say anything about how they relate
4. no you can not tell because the two are not linked together
5. No you can't determinewhether there is any significant difference between food preferences between differeing continents based solely on the data avaiable
6. No you can't tell because the choices from North America are so spread out and we can't tell what location the respondants were from
7. no you cannot make any inferences on if food choice differed based on continent
8. No you cannot tell whether food choices differed between respondents in different continents because these results are independent from each other
9. No, as there are no connections you can make between the data. You have no way to know joint distributions

Student responses

1. No, as they are, they show independent results. Like the graphs are independent. We need some kind of joint distribution.
2. No, because each graph shows choices from all respondents. You would have to separate the food choices by birth continent.
3. NO, because it only shows separate distribution between continents and foods.
4. No, because of marginal and joint distribution
5. No, because the first food choice graph does not show us the birth continent for each person.
6. No, because the graphs only show the overall food choices and birth continents, not the food choices within each birth region, so we would need a graph comparing food choices for Asia vs. North America.
7. No, because there is no connection between the two variables.
8. No, because these graphs show the responses among the whole population. To make any connection you need to relate the two tables together
9. No, because they do not show the variables together, so we can look at the food choice without knowing the continent, or the continent without knowing the food choice, which doesn't help for our question.

Student responses

1. No, because we can not see the relationship with different people.
2. No, because we can not tell which responses to both questions are associated with the other or not
3. No, because we can't see the relationship with the different answers in the two graphs
4. No, because we didn't match them together
5. No, because we don't know the relationship of the variables, we just know the variable statistics of the total sample population for each.
6. No, both graphs only include either continent information/food information, not both.
7. No, it shows both of them separately
8. No, since we are not looking at favorite foods among those born in Asia vs. NA. All we see is preferences for favorite foods and country of birth.
9. No, the first graph has nothing that would indicate information used in the second.

Student responses

1. No, the food choice graph has no info on place of birth
2. No, the graphs do not have a connection between them and can not be used to draw additional conclusions about our data
3. no, the graphs dont contain enough information to make a definitive conclusion
4. No, the results for birth continent aren't connected to favorite food.
5. No, these two graphs are standalone and don't tell anything about the other.
6. No, we cannot tell based on these graphs. We would have to put an interaction in to be able to tell whether food preference distribution differs between people born in different continents.
7. No, we don't know which birth_continent group replied what so we have to overlay the data in a different type of plot to see if that is true.
8. No, we don't know who belongs to what
9. No, we dont know marginal and joint distubtion

Student responses

1. NO, we need to check about the distribution between Asia, north america
2. No, we would want to show separate graphs for these. we would want to segregate based on whether respondents were born in asia or north america and then show the bar chart of food choices
3. No, you need to know which choices were associated with which respondants
4. No. Both graphs were based on the whole class and each class are technically independent, so we cannot tell how the food choices differed in each continent
5. No. We do not know the distribution of food choice for respondents born in each continent just by two separate distributions
6. No. We have to further filter out the Asians and Americans and check the distribution.
7. Not using just these graphs because we don't know the overlap, but we can use the data we have and figure this out by creating a new tab
8. The graphs don't tell us which students chose which favorite food. since the two variables are not complementing each other we don't know how the respondents responded to food based on their birth continent

Student responses

1. This graph cannot tell us anything about the relationship between the two categories, as each graph is in terms of its own category.
2. Using just these 2 I dont think so because the numbers dont show us the overlap
3. Using only these graphs, we cannot tell as both graphs do not overlap at all.
4. Using only these two graphs, you cannot make any inferences about if the two variables are related because you do not know which responses go together (or are from the same person).
5. We can not tell that food choices differed because they are separate pieces of information.
6. We can't yet tell just from the two graphs because we don't have the joint distribution.
7. We can't, because we don't know the set of responses that each survey respondent chose, aka knowing their birth continent and food preference at the same time. Therefore we cannot infer correlation.
8. We cannot because these graphs only include 1 variable but don't include all two of the food, birth continentq11
9. We cannot figure it out based on the graphs because we only have the marignal distribuions

Student responses

1. we cannot tell the relationship between food choices and the respondent's birth continent due to there is no linkage between the 2 tables
2. We cannot tell whether food choices differed between respondents just from these two graph as it doesn't show any correlation between them
3. We cannot tell whether food choices differed between respondents solely by looking at the two separate graphs. The data has been isolated for each so they are not related.
4. We cannot tell whether food choices were difference between respondents born in NA vs asia since there is no graph that is comparing them yet
5. we cannot yet becase we didnt organize by continent of birth.its generalized nd its combined between the multiple continents
6. We cannot, because there is no shown explicit correlation between birth continent and food type. The distributions cannot be compared.
7. We cannot. Bar graphs do not give any information on which respondent answered what.We would need to create a plot that relates the two

Student responses

1. We cant at face value, we havent seen which group of people majorirty voted for each food choice. we need to see a percentage breakdown per each country and each food choice to see
2. We cant tell yet because we are only given the amount of students from each place and the amount of students who like a particular food. We arent able to make a conclusion for te question because we dont have all the info.
3. We have the marginals, but by doing that we lose information about their joint distributions. Because of this, we cannot find this information from just the given plots.
4. Yes, people from different region may favor the food from their hometown.
5. You cannot tell whether food choices differed between Asian and North American respondents because the distribution and relation is not shown in the same graph.
6. You cant tell yet because you dont know the correlation and you cant see in either table which birth continent has more

Student responses

1. I feel like no because I am not sure if there is a correlation?
2. they cant tell us the why behind the choices
3. We can by seeing the correlation between food choices and countries of birth
4. Yes, because you can filter for respondents from North America and Asia, compute the food choices for those parties specifically, and compare
5. yes, then make a test
6. You could look at each continent and see differences between the choices of favorite food for both, but you would not be able to come to any conclusion at all.

Student responses

1. No
2. no because there are only 2 locations and many foods
3. No they are not related. There are more things to influence the output
4. typically asian originated foods would be preferred if someone was born in Asia, can't use marginal ditribution fo foods ab the joint ditribution

Student responses

1. Asia people might not like sushi and same way for the American and pizza
2. respondents born in Aisa tends to choose the food that is origin in Asia, and respondents born in America will choose variant kind of food.
3. Yes, because born in Asia tend to have median lean towards asian food and vise versa

Which foods appear most common in each birth-continent group? Describe one similarity and one difference between the two rows of counts.

99 anonymous responses

Sushi is most common in both groups	61
Responses stating sushi is the top choice for people born in Asia and in North America (often noting sushi dominates or is highest count in each row).	
Burritos much more common in North America	29
Responses highlighting burrito is popular or second-most in North America but uncommon or rare among those born in Asia.	
Other responses	3
Responses the model could not place reliably.	
Other responses	3

Response themes for question 9

Pizza appears only in North America

3

Responses noting pizza is chosen by North American-born respondents but not by Asian-born respondents (often used as a contrast).

Student responses

1. 10/18 people born in Asia chose sushi as their favorite food, and 26/80 people born in North America chose sushi. Both have sushi as their most common favorite food, but Asia sees sushi dominate with a much higher percentage
2. A similarity is that both groups have their favorite food being sushi. A difference is that foods like salad and pizza have only respondents from North America selecting them. It is interesting to see the high for Asian respondents being an option many north americans had but the north american top response having almost no asian people selecting it
3. Asian people have a stronger favorability towards sushi compared to the north americans, but there is still a decent majority for sushi in the north americans
4. Both continents have sushi as the most common favorite food but north america is more diverse and has more selections across the board while asia mostly has selections that are asian cuisines

Student responses

1. Both continents have Sushi as their most common favorite food, North America has a large number of votes for Pizza and Burritos, while Asia has one vote for burritos and none for Pizza
2. Both groups favored sushi, signifacantly more asians proportionately
3. Both groups had a relatively high proportion of people who liked sushi, and the North American born people tended to like burritos more often than the Asian people did.
4. Both groups prefer sushi the most, but north americans prefer burritos as well while asians don't.
5. Both had sushi number
6. Both people in Asia and NA both really like Sushi and have a majority there, but people from NA show an interest in burritos that does not exist in Asia.
7. Both Sushi, but North America tends to have people favoring pizza more.
8. Burrito is the second most popular food for those born in North America but not popular for those born in Asia. Sushi is most popular for both groups.
9. Everyone loves eating sushi, however the north america may have more variety of cusine options.

Student responses

1. For both Asia and North America, sushi is the most popular response. But for Asia nearly all people chose Sushi while for NA people have other choices
2. For those born in Asia, the most common food was Sushi. For those born in North America, Sushi was also the most common food, but Burrito ended up being more popular too (as well as other foods)
3. In both Asia and North America, sushi was the most common favorite food. However, burritos were much more favorited among students born in North America compared to those born in Asia.
4. In North America people like Pizza more than Asia. In both groups Sushi is a popular food.
5. In this class, folks from asia are more likely to have Sushi as their favorite food when compared to North Americas.
6. Individuals born in asia prefer sushi and pho over other foods by a dramatic margin, whereas those born in north america prefer sushi and burritos. Both prefer sushi, however each has a food the other doesn't seem to prefer.

Student responses

1. Most people chose Sushi for people born in Asia, whereas there's two favorite foods for people born in America: burritos and sushi. We can observe that survey respondents, regardless of continent, love sushi in general.
2. One similarity I saw between the rows is that a majority of people from each continent prefer sushi a their favorite food on the list, one difference was that the people born in Asia didnt mark pizza down as a favorite at all
3. One similarity is that Sushi is the food with most votes for both continent groups. One difference is that North America has a lot more people liking burrito than Asians
4. one similarity is that sushi was the most people across both one difference is that north america likes pizza burritos a lot more than asia
5. People from both continents seemed to like sushi alot while pizza was far more popular in North America
6. People in Asia and North America both like sushi the most however the % of people in Asia who like sushi the most is higher than the US

Student responses

1. Sushi and Pho are most common amongst students born in Asia. Pizza and burritos are more common in North America. When comparing both groups, sushi is the most common choice.
2. sushi appeared most often in north america and asia. however, burritos were also very popular in north america but not in asia, and pho was very popular in asia but not north america
3. sushi appears more common in both groups -- this is one similarity. The difference is that respondents born in Asia chose concentrately on sushi while respondents born in NA also chose other popular choices
4. Sushi appears most common among both but burritos is most popular in NA
5. Sushi appears most common among both continents, but burritos are more common in North America
6. Sushi appears most common in both Asia and North America. Pizza is more common in North America.
7. Sushi for the both continents. However, the percentage of people who love Sushi is different for the two continents.
8. Sushi in both is the most popular, but Burrito is a very close second for North America.

Student responses

1. Sushi is most common in both Asia and North America. However, this is a much more overwhelming majority in Asian respondents, while in North American respondents burritos are almost as popular as sushi (and pizza has 11 responses among NA respondents and 0 among Asian ones).
2. Sushi is most common in both groups. One difference is that North Americans prefer burritos way more than those born in Asia. One similarity is that barely anyone likes salad.
3. sushi is most common in both, but pho is the runner up in asia and burrito is the runner up in NA
4. Sushi is most common, burrito is the one different.
5. Sushi is popular among both continent groups, but pizza and burritos are mostly popular in North America
6. Sushi is still the most common food across birth continent groups, but the distribution of how much pho was chosen as a favorite food is higher in Asian birth countries
7. Sushi is the highest choice for both continents, but Asia has a higher proportion of its answer choices as sushi compared to North America
8. Sushi is the most common in both continents, pizza more heavily liked in North America

Student responses

1. Sushi is the most common in both groups. Although burrito is the 2nd most popular in North America's group, but not Asia's.
2. Sushi is the most common in both rows. Burritos are much more favored by north americans than asians. Something similar that I noticed is that both continents tend to enjoy asian food a lot.
3. Sushi is the most common in each birth-continent group. The difference is that there is more variance in the North America group, and a similarity is the liking for sushi
4. Sushi is the most common in each continent, however in north america pizza is a mildy common choice where no one selected in asia
5. Sushi is the most common off of pure quantity for both continents, but Asia had a less diverse favorite food spread compared to North America.
6. Sushi is the most popular in both continents, while North America has a larger variety of tastes instead of Asia, as Asia primarily favors Sushi heavily over all the other foods
7. sushi is the most popular in both group

Student responses

1. Sushi is the most popular in eah continent. Salad is considered the least popular food in boh continents. However, unlike burito was popular in America, it was not popular at all in Asia
2. sushi is the msot common one in both Asia and America.
3. sushi most common bc i see the biggest line
4. Sushi most common is both, but higher proportion in Asia
5. sushi was msot common in both groups, burrito was second most common in NA, but one of the least common in Asia
6. Sushi was the most chosen for both groups of people whereas burrito was chosen almost exclusively by people born in North America.
7. Sushi! Sushi is in both. I see more people answer burrito in north america, that is a difference
8. The most common food in each birth-continent group is sushi. One similarity is the lower number of people from each continent that like salad or stir-fry, and one different is that people born in north america seem to really like burritos, while only one person from asia said that was their favorite.

Student responses

1. The most common food in each continent is Sushi. A similarity is that sushi was extremely popular among both students from NA and Asia. A difference is that burrito was very popular among students from NA, but only received one vote from Asia.
2. The most common food is sushi in both continent groups. One similarity is that both continents liked sushi the most. One difference is that North America survey pool liked pizza more than Asian survey pool
3. The most common in both is sushi. One difference is that 11 people preferred pizza in North America and none in Asia
4. They all love sushi but north American love Burrito more
5. They both have sushi their food preference being the most common but Asia has a larger proportion who chose sushi

Student responses

1. America like Burrito, and Asia like suchi
2. Among people born in Asia, sushi was the most popular food choice. For people born in North America, burritos were the most popular.
3. Asia was sushi as most common, burrito was the most common in North America. One similarity is that they are whole numbers and one difference is that there is way more quantity in NA
4. both group love sushi, no asian love pizza,
5. both like sushi but people born in north america like burritos way more.
6. Burrito is popular in North America
7. burrito is very high for people born in the US and very low for people born in asia
8. burrito is very popular. in america compared to asia. both groups tend to like sushi
9. Burritos are really common in North America and not really in Asia and pizza is really common in North America and not really in Asia. Sushi has some commonality between both continents

Student responses

1. For Asia, pho and sushi, for North America burritos and sushi. Both groups had sushi as most common, but burritos was common for north America and ranked being some of the least common for Asia
2. Huge difference in Pizza and Brito for the two continents.ur
3. North American preferences are distributed among more food types more so than the food preferences for the Asia respondents
4. North amrican Brito, Asian Sushi
5. One similarity is both people from Asia and North America love sushi; One difference is that more people in North America tend to love Bonritto
6. One similarity is that they are both quite varied, while a difference is that pizza was only chosen by people born in North America.
7. People born in Asia mostly chose sushi as their favourite fod. On the other hand, in North America, people has a split choice between burrito and sushias their favorite food
8. People from North America and Asia really like sushi and people from Asia do like pizza as much as people from North America
9. Pizza seems much more common in North America than in Asia.

Student responses

1. pizza, sushi, Pizza appears in NA while sushi appears in Asia
2. Sushi appears most common in Asia and burritos appear most common in North America. The foods have a greater spread in North America versus Asia
3. Sushi appears most commonly and one similarity is the ration of curry and kebab to continent, and one difference is there is a lot more burrito responses in north america
4. sushi is most common in both asia and north america. burritos are the second most common in north america, but they are extremely uncommon in asia
5. sushi is the most common, burrito was close in north america. no matter the continent sushi was the most popular among the two asia/north america groups people in asia don't like burritos as much as people in north america
6. Sushi is the most popular food in Asia and North America. However, the second most popular in Asia is Pho while for North America it is the burrito.
7. sushi was most common in both groups and burrito was way more common in north america compared to asia

Student responses

1. Sushi was the most popular in both groups, but the groups differed on burritos, where almost no one from Asia preferred burritos, but almost as many Americans as liked sushi liked burritos.
2. the distribution for favorite foods for those born in asia was skewed towards sushi. For north americans, sushi still won, but at a much less extreme scale as burritos were close.
3. The proportion of those who like pho is a lot higher with respect to those from Asia than those in the U.S. In general, sushi is popular in both groups
4. They both have a high proportion of people who picked sushi. For people born in North America, however, a much higher proportion picked burrito

Student responses

1. Nobody born in Asia said Pizza was there favorite food, both c preferred sushiontinent
2. sushi was the most popular for both. in asia less people favored kebab and curry than north america
3. the curry as favorite food ratio is the same in each country. pizza is way different and burrito

Student responses

1. Both Asia and NA have sushi as their favourite food. but NA have more ditribution on burrito and other food.
2. Both people born in asia and usa prefer sushi the most. more people from the usa like pho than people born in asia
3. Sushi is most common in each birth-continent group. However, no one born in Asia selected pizza or salad, even though pizza was pretty popular relatively in North America.

Student responses

1. Pizza does not occur outside of North America as a favorite food
2. Sushi is most popular in both birth continents. For people born in Asia, no one said pizza was their favorite, so I think pizza might be a mainly American/European food
3. Sushi seems to be the most common for Asia, while that also holds true for North America. A difference is that Americans like Pizza and Salads more than Asia

Ten respondents reporting Asia chose sushi. Twenty-six of the 80 respondents reporting North America chose sushi. Does this mean sushi is more popular in North America? Explain.

0 anonymous responses

No responses were submitted.

Based on the percentages calculated above, is sushi more popular with Asian or North American students?

101 anonymous responses

Asian — higher conditional percentage	89
States sushi is more popular with Asian students because a larger proportion/percentage of Asian students chose sushi (mentions figures like ~55% vs ~32% or general percent comparison).	
North American more popular	6
States sushi is more popular with North American students (explicitly names North America).	
Total counts vs percentage nuance	5
Notes distinction between raw totals and conditional percentages or mentions sample size affecting interpretation (acknowledges totals might favor NA but percentages favor Asian).	
Other responses	1

Student responses

1. 55% of student in Asia chose sushi while only 36% of NA student chose sushi. So sushi is more popular in Asia
2. Asia
3. Asia
4. Asia
5. Asia, 0.56 compared to 0.32
6. asian
7. asian
8. asian
9. Asian

Student responses

1. Asian
2. Asian
3. Asian
4. Asian
5. Asian
6. Asian
7. Asian
8. Asian
9. Asian

Student responses

1. Asian
2. Asian
3. Asian
4. Asian (56%), North American was only about 32%
5. Asian because the percentage of asians who likes sushi was greater than the percentage of north americans who liked sushi
6. Asian since with small amount of asian, most of them prefer sushi.
7. Asian students
8. Asian students
9. Asian students

Student responses

1. Asian students
2. Asian students
3. Asian Students
4. Asian students.
5. Asian, because the proportion are bigger
6. Asian. Conditioning on continent of birth, the probability of choosing Sushi is much higher for Asian than NA students.
7. Asian. With 55% larger than 32%.
8. asians
9. based on the percentage, it is more popular with asian students

Student responses

1. Based on the percentages, sushi is more popular with Asian students.
2. I believe it is more popular with asian students
3. It is more popular with Asian students
4. It is more popular with Asian students because $0.56 > 0.32$
5. It is more popular with Asian students, NA students were more split on their favorite food
6. It should b more popular with Asian students
7. It would be more popular in asia in terms of a percentage of asian residents who picked sushi compared to north americans
8. It's more popular among Asian students
9. It's more popular among Asian students, since the proportion of Asian students who chose it as their favorite is higher

Student responses

1. its more popular with Asian students because 55% of asian respondents chose sushi while only 33% of north american respondents chose sushi.
2. More popular in Asian students
3. more popular with Asian students
4. More popular with asian students
5. More popular with Asian students because the percentage is much bigger
6. our surge data says asian prefer sushi more than north americans, as 56% of asian students preferred sushi compared to 32% of americans.
7. Sushi incorporates a larger portion of total responses among people born in Asia compared to people born in North Ameica
8. Sushi is far more popular with Asian students (~55%) than American Students (~33%).
9. sushi is more popular among asian students

Student responses

1. Sushi is more popular among Asian students
2. Sushi is more popular among Asian students as the percentage of asian students that selected sushi as their favorite food was around 55 and that percentage was only around 33 among north american
3. Sushi is more popular among Asian students.
4. Sushi is more popular among Asian students.
5. Sushi is more popular among Asians than North Americans since the probability distribution is much more skewed towards sushi for Asians than North Americans
6. Sushi is more popular amongs Asian students
7. Sushi is more popular in Asia as the proportion is higher for Asian.
8. Sushi is more popular with asain students because the condition probability given country is higher for asain students
9. Sushi is more popular with Asian (56%) than North American (32%) students

Student responses

1. sushi is more popular with asian students
2. sushi is more popular with asian students
3. Sushi is more popular with Asian students
4. Sushi is more popular with Asian Students
5. Sushi is more popular with Asian students (56% respondants chose sushi)
6. Sushi is more popular with Asian students according to the calculated percentages
7. Sushi is more popular with Asian students as it accounted for a higher percentage of the total respondents who were Asian.
8. Sushi is more popular with asian students as mor ethan hald f od asia studnets chose sushi while only 32% of norht america students chose sushi
9. Sushi is more popular with asian students because 56% of them picked sushi compared to north americas 32%

Student responses

1. Sushi is more popular with Asian students because a higher percent of students within the Asian group liked sushi compared to the North American group
2. Sushi is more popular with Asian students because the percent chance someone likes sushi and is asian is higher than the proportion of asian students so they disproportionately like sushi.
3. Sushi is more popular with Asian students, 55% of whom picked sushi as their favorite food, while only 32.5% of NA students picked the same.
4. sushi is more popular with asian students, as 56% of asian students favored it while only 32% of north american students favored it.
5. Sushi is more popular with Asian students, as a larger proportion of Asian students like sushi compared to students from North america
6. Sushi is more popular with Asian students, based on our sample size.
7. Sushi is more popular with asian students, because within each continent the percentage of students that prefer sushi is higher in asia.
8. Sushi is more popular with Asian students, looking at the proportion adding up to 1 for the columns

Student responses

1. Sushi is more popular with Asian students, since a larger proportion of Asian students chose Sushi as their favourite.
2. Sushi is more popular with Asian students, with 55.6% choosing sushi compared to 32.5% of North American students.
3. Sushi is more popular with Asian students.
4. Sushi is more popular with Asian students. You can tell because they have a higher percentage than north american students.
5. Sushi is more popular with Asians
6. sushi is morecommon among asian students based on data
7. Sushi is still more popular with Asian students because it had a 55% conditional probability that among students that like Sushi, there is 55% probability that the student is Asian.
8. This is more popular with asian students as compared to all the asian students, they have a larger percent liking sushi compared to north americans which still have it as its largest but it isnt as large as the asian percent
9. with Asian students

Student responses

1. For north american students
2. it is more common among north american students
3. Norb amerca
4. North america, the probability of continent given Sushi being the preferred food is higher for North America than Asia
5. North American students
6. Sushi is more popular with American students

Student responses

1. Although the North American students had a higher over total of students that liked sushi, the distribution of total Asian students skewed almost completely on the sushi side, so it is more popular with Asian students over North American students. The sample size was just larger than Asian students
2. Asian there are just more American students in the class
3. By raw percentage 33% of students from North America chose sushi while 50% of students chose sushi from Asia
4. From a total standpoint, it is more popular among NA students, but from a percentage standpoint, it is more popular among Asian students. I think percentage is a better way to track this question so I would say Asian students.
5. Sushi is more popular with Asian students as over half of Asian students selected this food. We want to use the margin equal to 1 to determine this question.

Using the row percentages, describe one difference in food choices between the two continent groups. Using the column percentages, describe one difference in reported continent among people choosing different foods.

0 anonymous responses

No responses were submitted.

What sort of statistical test could we use to settle this question?

0 anonymous responses

No responses were submitted.

Suppose sushi were equally popular in the two groups. Could random samples of 18 and 80 people produce percentages of 55.6% and 32.5%, or would a gap that large be surprising? Make a prediction before we run the test.

0 anonymous responses

No responses were submitted.

What does the test show us? Interpret the p-value. Can we conclude that sushi is more popular in one group than the other?

0 anonymous responses

No responses were submitted.

Suppose we want to repeat the 2×2 test for burrito versus not burrito. 1. What two outcomes form the columns of the table? 2. What numbers would you use for ``x`` and ``n`` in ``prop.test(x, n)``? 3. What equal-proportions assumption would the test begin with?

0 anonymous responses

No responses were submitted.

Questions asked during class

1. **Slide 1:** what is this slide about

This slide looks like the title or opening slide for the lecture. It introduces the main topic, which seems to be about studying sleep and eating habits using data. It does not give details yet about methods, variables, or results, so those would likely appear on later slides.
