

Data Clubs Post-Module Interview

Datasets used during the interview:

Census At School 2015 USA Interview Data n=200.CODAP: <http://bit.ly/dcinterview1>

Roller Coasters Type Year Max Height.CODAP: <http://bit.ly/dcinterview2>

Intro:

We're curious about your Data Clubs experience. We want to understand more about how kids your age think about data, including data that are new to you, and how you use tools like CODAP. Don't worry about right or wrong answers, we want to hear more about how you **think** about data. So, to learn more about your thinking, I'd like to interview you. Is that OK?

Part 1: Working with a novel database in CODAP.

Use the data set labeled: *Census At School 2015 USA Interview Data n=200.CODAP*. Have the CODAP file open in table view with the attributes in cm off screen so the student just sees those attributes in inches.

<https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=145822>

Making Sense of the CaS Dataset (3-4 minutes)

This data was collected from students in grades 4 through 12 in math classes that were doing data investigations. The students were from schools all over the United States. Take a moment to look at the data. Do you have any questions about anything you see? *[Answer any questions about what elements in the table mean or where the data came from that students might have].*

1. Let's look at this table together. We will focus on just the rows for a moment *[gesture across a row, making sure the student understands that the rows run horizontally across the table].*

Data scientists would refer to each row as a case. Looking at this first row, what can you tell me about this case?

When you look at all the data, what does each row in the table represent? *[Probe student thinking, does the student recognize that each row or case encodes data from one person?]*

2. Now let's look at the columns for a moment *[gesture up and down a column, making sure the student understands that the columns run vertically along the table].* Data scientists would refer to each column as an attribute. Looking at the attribute "Age_years" *[point to the column]*, what do the numbers in that column tell you?

What other attributes do you see in this table? Any others?

3. *Point to a cell in the table. What does this value right here tell you? [Probe to see if the student is thinking of a unique attribute value for a specific case.]*

Asking Questions and Exploring the CaS Dataset (7-8 minutes)

4. Let's think about the data we have in this table.

What questions could you investigate using this data?

Can you think of any others? Anything else?

5. Choose one question [*this should be a student-generated question*] that you could investigate using this database. How would you approach answering your question using the tools in CODAP? Show me on the screen. Talk me through what you are doing. [*If the student can't remember how CODAP works say: If you can't remember how to do something in CODAP, you can tell me what you want to do, and we can figure it out together.*]

6. *After student has created a visualization to help him or her make sense of the data ask:*

- Can you explain the graph to me?
- What do you notice when you look at the data shown this way?
- Do you see any patterns? Explain what you are seeing.
- Do you feel like you were able to answer your question? Why or why not?

Can the Question be Investigated with the CaS Dataset? (15 minutes)

7. Next we are going to look at some questions someone else came up with. For each question, think about whether you could answer it with this data or not. Remember, your job is to evaluate the fit between the question and this data—there may be questions that you'd need different or additional data to answer. [NOTE TO INTERVIEWER—IF ONE OF THE QUESTIONS BELOW WAS ADDRESSED ABOVE, JUST HAVE THE STUDENT THINK BACK TO THE EARLIER WORK, RATHER THAN GOING THROUGH THE GRAPHING PROCESS AGAIN.]

- **Are younger teens more likely to take a bus to school than older teens?**

If not: Why not? What information would you need that isn't in the dataset? How might someone get that information?

If yes: Can you show me how you'd explore the question using CODAP? [*If the student can't remember how CODAP works say: If you can't remember how to do something in CODAP, you can tell me what you want to do, and we can figure it out together.*]

- **Do more males than females drive themselves to school?**

If not: Why not? What information would you need that isn't in the dataset? How might someone get that information?

If yes: Can you show me how you'd explore the question using CODAP? [*If the student can't remember how CODAP works say: If you can't remember how to do something in CODAP, you can tell me what you want to do, and we can figure it out together.*]

- **Are students with longer arm spans more likely to play basketball?**

If not: Why not? What information would you need that isn't in the dataset? How might someone get that information?

If yes: Can you show me how you'd explore the question using CODAP? [*If the student can't remember how CODAP works say: If you can't remember how to do something in CODAP, you can tell me what you want to do, and we can figure it out together.*]

- **Is travel time to get to school related to the size of the town or city that a student lives in?**

If not: Why not? What information would you need that isn't in the dataset? How might someone get that information?

If yes: Can you show me how you'd explore the question using CODAP? [*If the student can't remember how CODAP works say: If you can't remember how to do something in CODAP, you can tell me what you want to do, and we can figure it out together.*]

- **Is travel time to get to school related to how the student travels to school (walk, bus, car, etc.)?**

If not: Why not? What information would you need that isn't in the dataset? How might someone get that information?

If yes: Can you show me how you'd explore the question using CODAP? [*If the student can't remember how CODAP works say: If you can't remember how to do something in CODAP, you can tell me what you want to do, and we can figure it out together.*]

Part 2: Working with a novel database in CODAP.

Use the Roller Coaster Time Type Speed Height Dataset. Have it opened in table mode with the box explaining the data minimized.

<https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=145866>

Making Sense of the Roller Coaster Dataset (3-4 minutes)

This data set has information on 157 roller coasters in the United States that opened between 1915 and 2016. If I opened up the text box over here (right corner of the screen) it would give you a whole bunch of information about the data. However, because we don't have a ton of time to explore everything, the table you are looking at has just 3 of the attributes (there were 16 in the original dataset). The second column here shows the attribute "Type"—it indicates the material of the track—either wood or steel.

8. What other attributes do you see? *[Ask: Any others? if not all have been mentioned]*
 What information does each attribute give you? *[If he or she is not sure here are the descriptions--“Year Opened” tells you the year the roller coaster was opened. “Max Height” is the highest point of the roller coaster in feet.]*

9. Next let’s look at the rows in the table. Each row in the table is a case. The index down the side here numbers the cases. What do you think each case represents? *[If a student gives another answer, have them consider whether each case might be a roller coaster].*

Asking Questions and Exploring the Roller Coaster Dataset (10 minutes)

10. Let’s zoom out and think about the data we have available in this table.
 What questions could you investigate using this data? Can you think of any others?
If the student has not generated a question involving “Year Opened” and another attribute, ask: I’m curious about the attribute, “Year Opened”. Are there other attributes in this dataset that you think might be related to Year Opened? Can you come up with a question that involves Year Opened and at least one other attribute?

11. *Focus on the question the student posed having to do with Year Opened.*
 How would you investigate this question using CODAP?

Have the student explore his or her question with CODAP. If a student forgets how to access some of the CODAP functions he or she can explain the move to the interviewer and the interviewer can help the student make the moves he or she intends.

12. Can you explain your graph to me?

Probe as appropriate:

- Can you explain your graph to me?
- What does each dot represent?
- What do your axes show?
- What does the color tell me?
- What do you notice when you look at the data shown this way?
- Do you see any patterns? Explain what you are looking at.
- If you were a reporter using this graph in an article, what might your headline read?

Part 3: Thinking about Data Science in General (5 minutes)

Now we are going to shift gears and I want to ask you some general questions about what you have learned and how you view data science.

1. What is data science and what do Data Scientists do? (Probe: Anything else?)
2. What did you like most about this experience? Do you feel like you look at data science any differently now than you did before?