API CAN CODE

COLLEGE OF EDUCATION

Unit 3: Data Science Practices

Lesson 3.4: Graphs and Figures for Two or More Variables

In this lesson, students will learn about different multivariate graphs, namely the scatter plot and binned dotplot.

Duration: 90 minutes

Objective: By the end of this lesson, students will know how to create scatter plots and binned dotplots in CODAP and be able to interpret them.

Lesson Walkthrough Video: Unit 3 Lesson 4 - Teacher Walkthrough

CSTA Data Science Specialty Standards in this Lesson

Identifier	Concept	Subconcept	Standards
S1-DSC-MI-07	Data Science	Interpreting Models, Problems, & Results	Interpret the results of data analyses to explain patterns, anomalies, and trends, and connect them back to the original problem or research question.
S1-DSC-VZ-10	Data Science	Visualization	Create visualizations that communicate key findings to diverse audiences using appropriate chart types and formatting.

Lesson activities

Warmup (10 min)

(CSTA standards in this activity: 3A-DA-11)

- From October 26, 2023: What matters when choosing a college?, Consider the following questions:
  - What kind of graph is this?
  - What variable(s) do you see represented? Is there more than one?
  - What do you notice? What do you wonder?
  - What story does this graph portray?

Roller Coasters (60 min)

(CSTA standards in this activity: 3A-DA-11, 3A-DA-12)

- Open the Roller Coasters built-in dataset in CODAP

This work was made possible through generous support from the National Science Foundation (Award # 2141655). API Can Code is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License

- Examine the data. Feel free to create any graphs that help you make sense of the dataset. (10 min)
  - What kind of data is this?
  - What's going on in this dataset?
  - What do you notice? What do you wonder?
- Create a graph with Max Height on the x-axis and Top Speed on the y-axis. This creates a **scatter plot.** (5 min)
  - How would you describe the relationship between these variables based on the graph?
- Describe specific characteristics of the graph. (5 min)
  - What is the **direction** (positive or negative)?
  - What is the **form** (does it appear to be mostly linear, or non-linear)?
  - How **strong** is the relationship (is the association clear, or cloudy)?
  - Do there seem to be any major **outliers?**
- Review the slides on Direction, Form, Strength, and Outliers with students (10 min).
- Adding a third variable (5 min)
  - Drag the Drop, Length, Type, Year Opened, or Age Group variable onto the *middle* of the graph to color the points. What do you notice?
- Click the Ruler icon and add a Least Squares Line to the graph. (5 min)
  - **Teacher:** This line minimizes squared vertical distance, or squared **residual** distance, to all the points in the graph. You can see these squared residuals by going back to the Ruler icon and adding "Squares of Residuals"
  - The equation of the Least Squares Line (aka **line of best fit** or **regression line**) includes two numbers: 0.193 and 33.4. How do we interpret each of these numbers? (Hint: Which is the slope? Which is the intercept?)
  - Does the slope make sense? Does the intercept make sense?
- Now, create a new graph with "Age Group" on the x-axis and "Type" on the y-axis. Note that this doesn't create a scatter plot, but rather a **two-way binned dot plot.** (10 min)
  - What's different about the variables used to create this plot? Why do the dots get organized into groups instead of scattered across the graph?
  - Click on the Ruler icon, then Show… Percent, and finally select the Row button. What does the 34% in the top right box mean?
  - Now, go back to the Ruler icon, leave Show… Percent selected, but select "Column" instead of Row. What did the number in the top right box become? Why did it change?

This work was made possible through generous support from the National Science Foundation (Award # 2141655). API Can Code is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License

        - Based on this data, do you think amusement parks are increasingly building steel roller coasters instead of wooden ones? What evidence do you have?
    - Create a new graph, with Age Group on the x-axis. This time, drag Type onto the *middle* of the plot, which should color the points. Click the Ruler icon, and check "Fuse Dots into Bars." Then click the Ruler icon again, and select Scale… Percent. (10 min)
        - What information is this graph presenting? Does this make it easier to answer the question we just asked about steel vs. wooden coasters over time? How does this graph show that?

Scatter Plots (10 min)

- Students are given scatter plots of NFL trends, Stephen Curry, and the relationship between GDP and life expectancy.
- Discuss with the students about these graphs and ask them:
    - What can you tell about this graph? (Consider its direction, form, strength, and whether there are outliers.)

Exit Ticket (10 min)

(CSTA standards in this activity: 3A-DA-12)

- Students are given a scatter plot of test scores and final grades.
    - Describe the direction, form, and strength of this scatter plot.
        - (positive), (roughly linear), (moderate or weak), and (there is a point around (90, 60) that students will likely identify as not fitting the pattern)
- Students are given a table of pre-med major percentages by class year. Do you think there is a relationship between pre-med major and year? Why or why not?
    - We want students to say: yes, there is a relationship, because the percentages *don't* match (suggesting that fewer upperclassmen are pre-med majors, so there is a relationship there)

Assessment:

Assess student understanding through participation in class discussions and class activities.

This work was made possible through generous support from the National Science Foundation (Award # 2141655). API Can Code is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License