<p style="text-align:center">**Activity:** *Uncovering Educational Disparities*</p>

Introduction

How important is it to you that your country's educational system is good and fair to everyone? How can you tell if your country is doing a good job educating its students? Or if a particular region of your country is doing a worse job than other regions? Regarding fairness, how can you tell if educational results differ for different subgroups of the population?
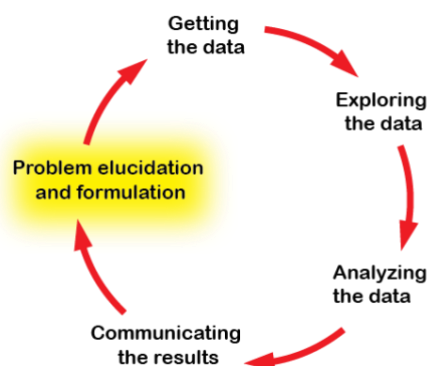

Figure 1a: *The basic cycle of learning from data*. Adapted from [9], p. 8. Copyright 2019 by The International Data Science in Schools Project.

Let's look at some data on the education system in the United States. Maybe the questions we ask about this data can give you ideas about analyzing similar data in your own region or country.

We'll be keeping track of what part of the Cycle of Learning from Data we're in as we go along, and we'll see that in some exploratory studies you don't have to follow its order exactly. It's mainly there to remind us that each of its aspects is important. ***As you move through this activity, highlight which part of the cycle diagram applies to what you are doing.***

The data set you will be using has education-related data for 49 of the 50 states in the United States (Hawaii couldn't be included because it doesn't generate data in a way comparable to the other 49 states). The data has been assembled from a variety of government databases that attempt to be a complete census of students in each state, rather than a random sample.

Use the following link to access a CODAP file, EdDataUSAstates, where we have collected education-related data (school year 2013-2014): https://bit.ly/EdDataUSAstates

Save your own version of this file on your computer or Google Drive. (This way, you can keep any graphs that you find interesting and may want to refer back to later.)

Starting to Explore the Data Set

When we're exploring a data set, it's often helpful to think of one particular data point to begin with. Since this data set is organized by state, pick a state to keep in mind; if you live outside the United States, maybe pick one you've heard of, or are curious about, or have friends or family who have visited there or lived there. Call that state "your state" for now.

Question 1) How much do teachers get paid in the United States?

> *a.* Create a one-variable graph with average adjusted teacher salary (Adj_TchSal) on the *x*-axis.

> *b.* Now scroll through the table to find your state. Click on your state's row in the table.

> *c.* Note that the data point corresponding to your state is now highlighted on the graph. Try it the other way. Click on an interesting point in the graph, and notice what happens in the table.

> *d.* You can delete this graph by clicking on the "X" in the top right corner.

Question 2) Individually or in pairs, take a few minutes to look at the other variables in this data set (you can see their definitions by hovering over each attribute name or looking at the "Attribute Descriptions" text box). What sparks your curiosity?

Write down a statistical question that you can answer using some of this data.

Question 3) Once your whole class has agreed on one question, create one or more graphs to analyze the relevant data. Decide which way the data can be best displayed to answer the question.

Question 4) Write down your interpretation of the results.

Question 5) What have you recently heard in the media about educational fairness? Briefly share about this with your neighbor. Note here any possible connections between things you and your peers have heard in the media about educational fairness and the data found in the EdDataUSAstates data set.

Examining Graduation Rates: Graphing Quantitative Variables

Next you'll go through some questions that the authors explored using this data set. Our first set of questions relates to statewide high school graduation rates.

While it's not the only way to measure the health of an education system, a state's public high school graduation rate is one such indicator. Graduation rates by state were retrieved from the National Center for Education Statistics [18] and included in the EdDataUSAstates data set.

Question 6) The definition of public high school graduation rate in this data set is: the number of students in the group (by income or race or overall) who graduate in 4 years with a regular high school diploma, divided by the number of 9th grade students in the group that started high school (after adjusting for issues like transfers in and out during those 4 years). What are some strengths and weaknesses of this definition?

Question 7) How do low-income graduation rates compare to overall graduation rates?

    a. First make two 1-variable graphs, one for GradRate_All and one for GradRate_LI. Use the ruler icon to add a box plot to each graph. Adjust the scale on the *x*-axes so that the two graphs can be compared.

    b. Compare the following features of the graphs:

        i. **S**hape (skewed, flat/uniform, mound-shaped, etc.)

        ii. any **O**utliers

        iii. **C**enter

        iv. **S**pread, or variability (i.e. range, maximum and minimum values, interquartile range)

c. Interpret these results, referring to all of the above features. Explain in detail what story this data is telling in answering our question of how low-income graduation rates compare to overall graduation rates.

The type of analysis you performed can be remembered using the acronym **SOCS**: **S**hape, **O**utliers, **C**enter, **S**pread(/variability).

Question 8) What other questions do you have about low-income graduation rates compared to overall graduation rates that cannot be answered using these 1-variable graphs?
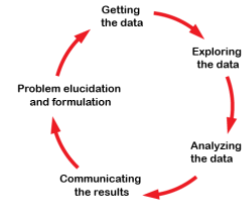
Question 9) Create a 2-variable graph to further investigate the relationship between low-income and overall graduation rates.

a. Interpret this graph in context.

b. Using the ruler icon, add the plotted function $f(x) = x$ to your graph. (In the dialog box, you will only need to type in the letter x.) What purpose can this line serve in understanding the relationship between low-income and overall graduation rates (our statistical question)?

c. Does this new graph help to answer any of the additional questions you listed in Question 8?

d. To get even more information from this graph, use the ruler icon to add a least-squares line to your graph.

What statement could you make about data points above the line? Below the line?

Are inequities in graduation rates worse in states that have more poverty overall? One hypothesis could be that high state-wide poverty levels mean there is not enough education funding to help poor school districts very much.

Question 10) Without looking at any data, take a guess about any relationship between a state's graduation rates and its overall poverty rate:

a. Overall graduation rate and state poverty rate?

b. Low-income graduation rate and state poverty rate?

c. Graduation rate inequity (between low-income and overall) and state poverty rate?

Now create a new attribute to represent graduation rate inequity, or the difference between low-income and overall graduation rates:

Step 1: In CODAP, use the plus sign in the top right of the case table to create a new attribute, GradRate_LI-All, to represent graduation rate inequity.

Step 2: Click on the attribute and use "edit formula" to make GradRate_LI-All equal to GradRate_LI - GradRate_All.

Question 11) Without consulting your other group members, create graphs in CODAP to investigate your guesses from Question 10.

a. Sketch the graphs that you created. (What attributes were on the axes?)

b. Now share your graphs with your group. Did you make the same graphs? If you didn't make the same graphs to answer a question, is one of them more effective at telling the data's story on that question? Why or why not?

c. What did you discover in this investigation? Write down your interpretation of the graphs, describing any relationships you see between state poverty rates, overall graduation rates, low-income graduation rates, and graduation rate inequities.

d. Discuss your thoughts about your results. Generate at least one hypothesis that might explain your findings.

Examining Racial Segregation: Graphing Categorical Variables

Our next question has to do with racial segregation in schools: are some states' schools more racially segregated than others? If so, is this usually true for one racial group or does it point to a more general problem?

The EdDataUSAstates data set includes data on racial segregation in schooling systems [20]. The attributes HighSeg_B and HighSeg_H show whether or not a state's school districts are highly segregated by race.

Question 12) In this data set, a state is listed as being highly segregated for Black students if at least 40% of Black students in the state attend schools with at least 90% Black students, and similarly for Hispanic students.

a. In your own words, interpret this definition of a highly segregated state.

b. What are some strengths and weaknesses of this definition?

Next, we will analyze the data on segregation:

Step 1: In CODAP, create a graph with HighSeg_H on the *x*-axis and HighSeg_B on the *y*-axis. You have created a **two-way binned plot**, one way of displaying bivariate categorical data.

Step 2: Use the ruler icon to display Count.

Question 13) Interpret the graph in your own words, as if you were explaining it to a friend.

Now let's try another way of analyzing/graphing the data.

Step 3: Now create another graph with HighSeg_H on the x-axis. This time, drag HighSeg_B to the *middle* of the graph. This will color the dots according to that attribute.

Step 4: Notice that a bar graph icon has appeared on the sidebar of the graph. Use that to select "fuse dots into bars."

Step 5: Use the bar graph icon again to change the scale from Count to Percent. You have created a **segmented bar graph**. If you hover over different parts of the bars, CODAP gives you information about the percentages within each bar. (You can even click on a segment to highlight all of the states in that category in the table.)

Question 14) Does there seem to be any relationship between racial segregation for Black students and Hispanic students? Explain your answer by referring to specific features of the segmented bar graph.

Examining Economic Segregation: Combining Categorical and Quantitative Variables

Is there any relationship between racial segregation and economic segregation among school children?

Investigate this using the attribute "EconSeg." This attribute reflects the degree to which children attend schools with high levels of wealth or poverty compared to neighboring schools [25].

Question 15) Create a new graph with EconSeg on the *x*-axis and HighSeg_B on the *y*-axis. Use the ruler icon to add the mean for each group.

What conclusion might you draw from this graph?

This data set has data from almost all of the 50 states. However, statistical studies can often only get data from a smaller group of items. Let's pretend that's our situation and see what happens. Choose one of the Regions in the data set (whichever Region you like), and we will suppose we had data only from that Region.

Drag the attribute "Region" to the middle of the graph to color the dots by that attribute. Click on the legend square corresponding to *your* region to highlight all of the states in your region. Then use the eye icon to hide all unselected points and show only the points in your region.

Question 16) If you had data from only your region, would it make your argument stronger or weaker than using data from the entire United States? Explain.
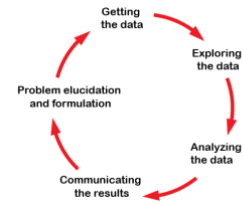
Our next question is about federal spending on education. Do all high-poverty schools get the same amount of federal revenue per pupil? If not, what are some factors associated with high levels of federal aid?

Question 17) Create a new graph with FedPPR_Pov (federal spending on high-poverty school districts) on the *x*-axis. On the *y*-axis, plot the attribute HighNativePop (the states with the highest percentages of Native Americans have a "yes" value). Use the ruler icon to add a box plot. Interpret this graph, using the SOCS mnemonic to help you attend to all of the relevant features.

While Native Americans experience high rates of poverty, they have a strong history of self-advocacy around education issues [15]. Their accomplishments can serve to inspire students and community leaders from other marginalized groups to be active in self-advocacy efforts.

Question 18) Suppose you wanted to present an argument that areas with a high Native American population receive more federal funding for high-poverty districts than areas without a high Native American population. If you used a graph like the one in Question 17, what features would you highlight to bolster your argument?

Question 19) Now that you've done all of this analysis, what further questions do you have about education and fairness that might need data beyond this data set?

[*End of activity*.]