STATISTICS TEACHER
SUPPORTING THE TEACHING AND LEARNING OF STATISTICS
STatistics Education Web STEW
online journal of K-12 statistics lesson plans

Lizards Task Levels B and C

Teacher Version with Student Version

Authors: Kaycie Maddox (American Statistical Association); Christine Franklin (University of Georgia Emeritus Professor of Statistics)

Lesson Overview: This lesson provides a secondary data set for students to explore to develop understanding of classification, a data science concept, using accessible technology, CODAP. Students will also be engaged with answering statistical questions utilized at each of the four phases of the Framework for Statistical Reasoning. The use of secondary data on the habitats of lizards combined with posing questions will strengthen the opportunity for students to think and reason statistically. For Level B, students will utilize boxplots with one quantitative variable and two categorical variables to interpret misclassification rates to determine classification. For Level C, students will use scatterplots with two quantitative variables and two categorical variables (multivariable) to interpret linear models and determine classification.

Types of Data:

- Level B: one quantitative variable and two categorical variables.
- Level C: two quantitative variables and two categorical variables.

Learning Objectives:

Common Core State Standards Mathematics	Next Generation Science Standards
CCSS.Math.Content.6.SP.B.4 Display numerical data in plots on a number line, including dot plots, histograms, and box plots. **CCSS.Math.Content.7.SP.A.1** Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences. **CCSS.Math.Content.8.SP.A.1** Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association. **CCSS.Math.Content.8.SP.A.3** Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.	**MS-ESS3-3 Earth and Human Activity** Apply scientific principles to design a method for monitoring and minimizing a human impact on the environment.* **ESS3.C: Human Impacts on Earth Systems** ● Human activities have significantly altered the biosphere, sometimes damaging or destroying natural habitats and causing the extinction of other species. But changes to Earth's environments can have different impacts (negative and positive) for different living things. (MS-ESS3-3) ● Typically as human populations and per-capita consumption of natural resources increase, so do the negative impacts on Earth unless the activities and technologies involved are engineered otherwise. (MS-ESS3-3),(MS-ESS3-4) **HS-LS2-2 Ecosystems: Interactions, Energy, and Dynamics** Use mathematical representations to support and revise explanations based on evidence about factors affecting biodiversity and populations in ecosystems of different scales. **HS-LS2-6 Ecosystems: Interactions, Energy, and Dynamics** Evaluate the claims, evidence, and reasoning that the complex interactions in ecosystems maintain relatively consistent numbers and types of organisms in stable conditions, but changing conditions may result in a new

CCSS.Math.Content.HSS.ID.C.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.	ecosystem.

Audience: Level B: 7th grade students, 8th grade students; Level C: Algebra I students

- This lesson has been tested with middle and high school students as well as with pre-service and inservice teachers of middle and high school.
- Prerequisites: Prior to this lesson, students should have had experience formulating and interpreting statistical investigative questions as well as creating and interpreting boxplots.
- It is not necessary for teachers or students to access the link to the Lizards data set in CODAP to complete this lesson. The data distributions from the data set and CODAP are provided for exploration within the lesson.
- This is the link to the **Lizards data set** within CODAP. It can be explored by teachers and students for the Extension Lesson following the Lizards Lesson.

Time Required: 2 days (50-60 minute sessions)

Lesson Plan

ENGAGE

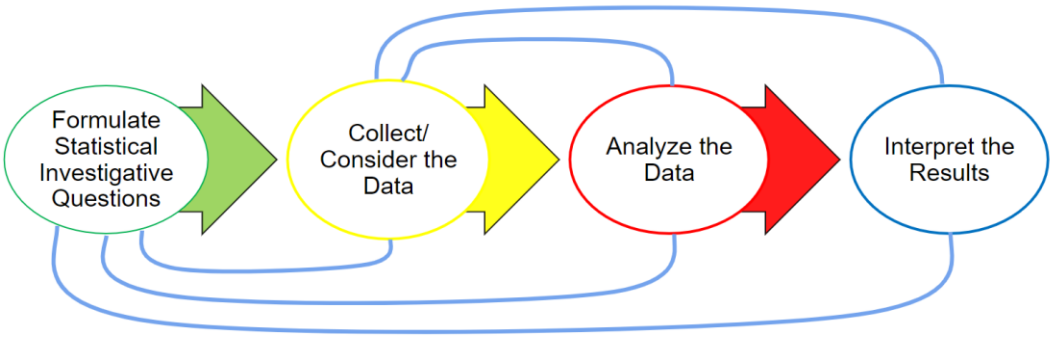
Students in a biology science class have been studying about disturbed habitats (those that have human development) and natural habitats for lizards. They are interested in knowing if the mass of a lizard is useful in predicting the type of habitat a lizard lives.

Notice and Wonder:
Ask students to work in partner pairs and write down three things they notice about this scenario and three things they wonder about it. This reading strategy will provide them with the opportunity to read the scenario closely and make sense of it themselves. There are several vocabulary words in this brief situation, so be sure to call attention to them and define/clarify them with students in the event the students do not ask about them as part of what they wonder.

Also, ask students to write their own statistical investigative question for this scenario. Share these with the class through Padlet or another technology app or with individual whiteboards.

EXPLORE: Level B

As standard procedure for the tasks used in K-12 classrooms, students will be utilizing the Statistical Problem-Solving Process to guide their thinking and understanding.



Ask

The students are interested in exploring the following statistical investigative question: **How can a lizard's mass be used to predict whether it came from a "Natural" or a "Disturbed" habitat?**

Students will answer the following interrogative questions to better understand the statistical investigative question.

Interrogative Questions:

- Is it a specific type of lizard or any lizard?
- Is it for a specific geographic region?
- What is meant by *mass*?

Utilize the Think-Pair-Share strategy with student partner pairs to initially discuss these interrogative questions. Then call on various groups to share their answers with the entire class. These questions should help to clarify the statistical investigative question, though the students may not yet have enough information to answer them all.

Consider building a **TIP Chart** for the vocabulary words as they come up in the lesson. Fill in definitions or information in that column, remembering that more information can be added later. The picture column can be an example, a picture, or a notation or formula. The TIP Chart should be large and on display in the classroom and can also be kept in student notebooks.

Term	Information	Picture
Natural habitat		
Disturbed habitat		
mass		
Secondary data set		
Primary data set		
. . .		

Collect

The students use a data set where a biologist captured (randomly) a number of individual lizards across two habitats on four islands in the Bahamas. The biologist took various measurements including mass, head depth, and hindlimb length. The students are considering a *secondary data set* – data collected by another researcher in the Bahamas. If the students had collected the data, the data would classify as *primary data*.

81 lizards in the study came from natural habitats.

79 lizards in the study came from disturbed habitats.

Students will answer the following interrogative questions to better understand the data collected.

Interrogative Questions:

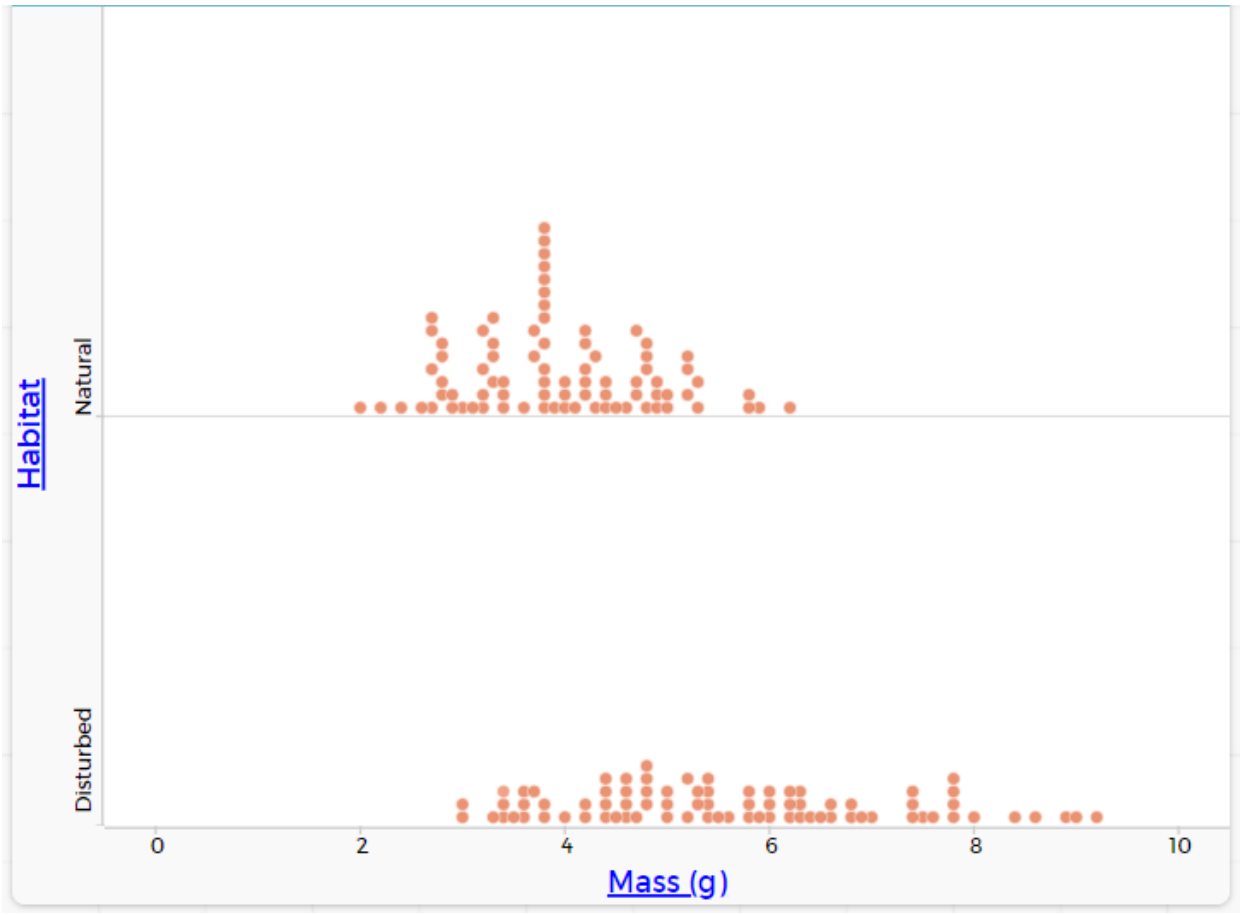
- What type of lizards were used in the study?
- What are specific details about the Bahamas geography? How are the four islands similar/different?
- How were mass, head depth, and hindlimb length measured? What units were used?
- What was the reason the lizards were randomly selected?
- Can this secondary data set reasonably answer our statistical investigative question?

Utilize the Think-Pair-Share strategy with student partner pairs to initially discuss these interrogative questions. Then call on various groups to share their answers with the entire class. These questions should help to clarify the statistical investigative question, though the students may not yet have enough information to answer them all.

EXPLAIN and ELABORATE

Analyze

Here is a dotplot representation of the distributions using mass of the lizards and the habitats of the lizards collected in the biologists' study. Students are asked to notice three things about the dotplot representation and wonder three things. This strategy is utilized again so that students will look closely at the representation and begin the process of analyzing the data presented there. ENCOURAGE STUDENTS TO USE COMPARISON LANGUAGE IN ADDITION TO DETAILS ABOUT EACH TYPE OF LIZARD. LOOK FOR OVERLAPS, GAPS, AND CLUSTERS.

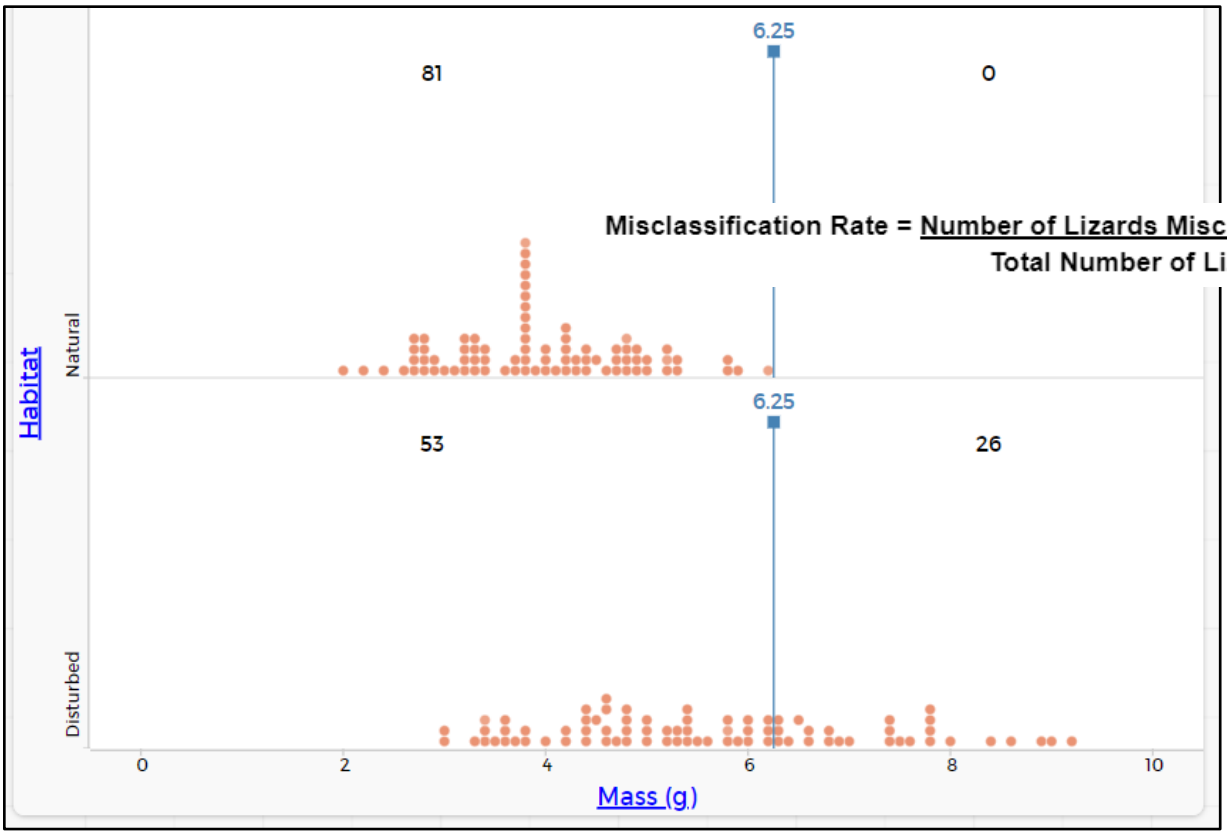


Students are looking at this known data (secondary data) about masses and habitats from captured lizards so they can be able to answer the statistical investigative question: **How can a lizard's mass be used to predict whether it came from a "Natural" or a "Disturbed" habitat?** This means that in future experiments, they can make predictions about the type of habitat a lizard may have come from just by knowing its mass, something measurable.

Students will be asked, **"Where would you draw a vertical cutoff line on the dotplot representation to classify lizards as having come from a natural habitat?"** Allow time for the partner pairs of students to explore the location for the cutoff line. Students can utilize a pen or pencil to serve as their vertical line for counting the dots in each of the four quadrants. Have students share their reasoning.

After students have had time to suggest a cutoff line, explore the trial rule below with the students

Trial Rule #1: Classify all lizards with a mass less than 6.25g as natural. Classify all lizards with mass less than 6.25g as natural to the two distributions, since there is obvious separation between the mass of the lizards in the natural and disturbed habitats at 6.25 grams.



What do the numbers (81, 0, etc.) on the dotplot representation mean?

81 *means 81 lizards have weights less than 6.25 g and are, therefore, classified as natural AND actually came from a natural habitat.*

0 *means that 0 lizards have weights more than 6.25 g and are, therefore, classified as disturbed AND actually came from a natural habitat. -* ***MISCLASSIFIED***

53 *means that 53 lizards have weights less than 6.25 g and are, therefore, classified as natural AND actually came from a disturbed habitat. -* ***MISCLASSIFIED***

26 *means that 26 lizards have weights more than 6.25 grams and are, therefore, classified as disturbed and actually came from a disturbed habitat.*

But now we ask the analysis question, "How often do we misclassify the lizards based on using this classification rule?"

	Truly from natural habitat	Truly from disturbed habitat
Classified as "natural"	81	53
Classified as "disturbed"	0	26

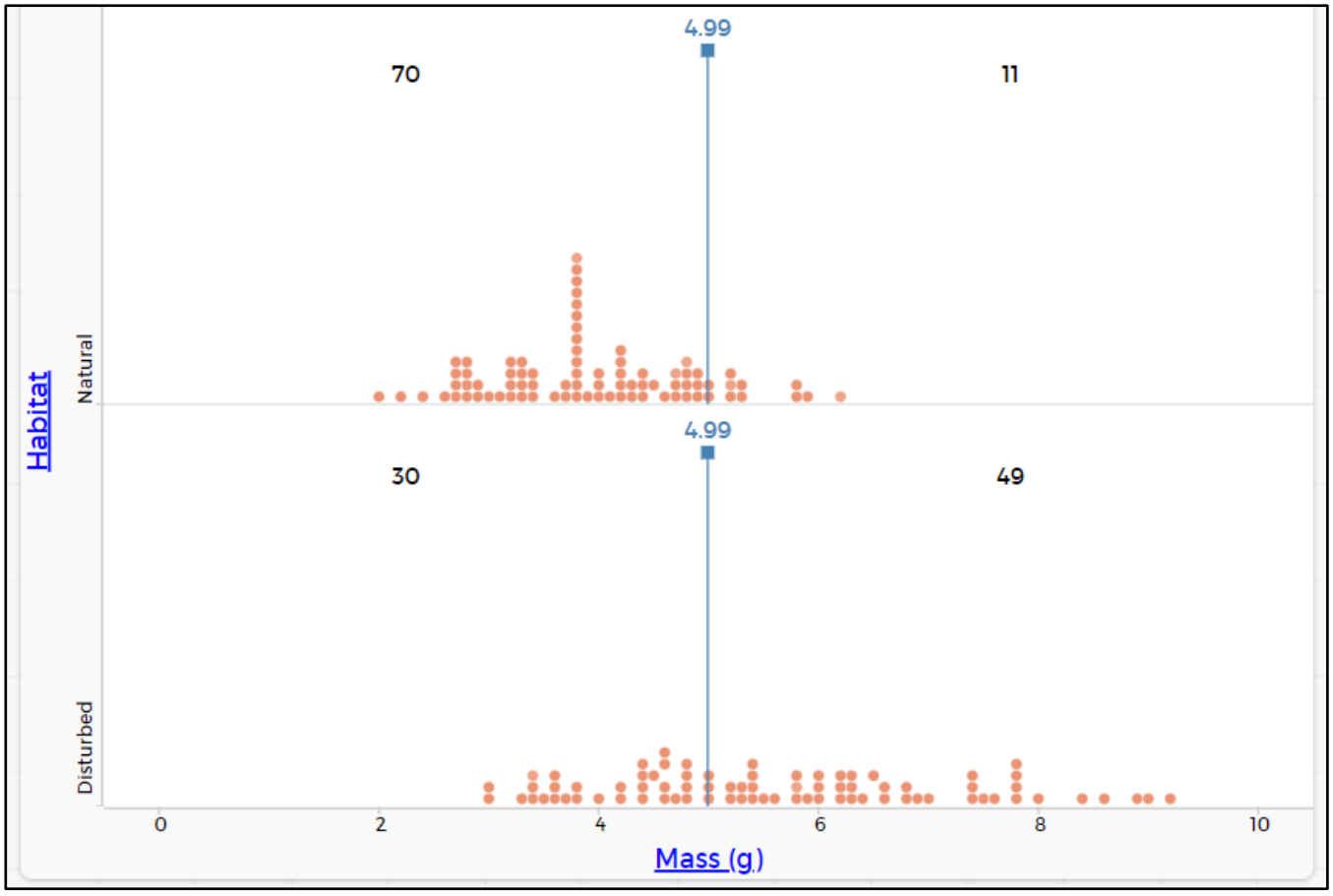
What is the misclassification rate for our vertical line drawn at 6.25 grams?

$$\text{Misclassification Rate} = \frac{53}{160} = 33\%$$

[Type here]

What does this misclassification rate tell us?
33% of the lizards captured would be misclassified as natural if using the rule that the mass is less than 6.25 grams to classify them as from a nature habitat.

Trial Rule #2: Classify all lizards with a mass less than 4.99 g as natural.



	Truly from natural habitat	Truly from disturbed habitat
Classified as "natural"	70	30
Classified as "disturbed"	11	49

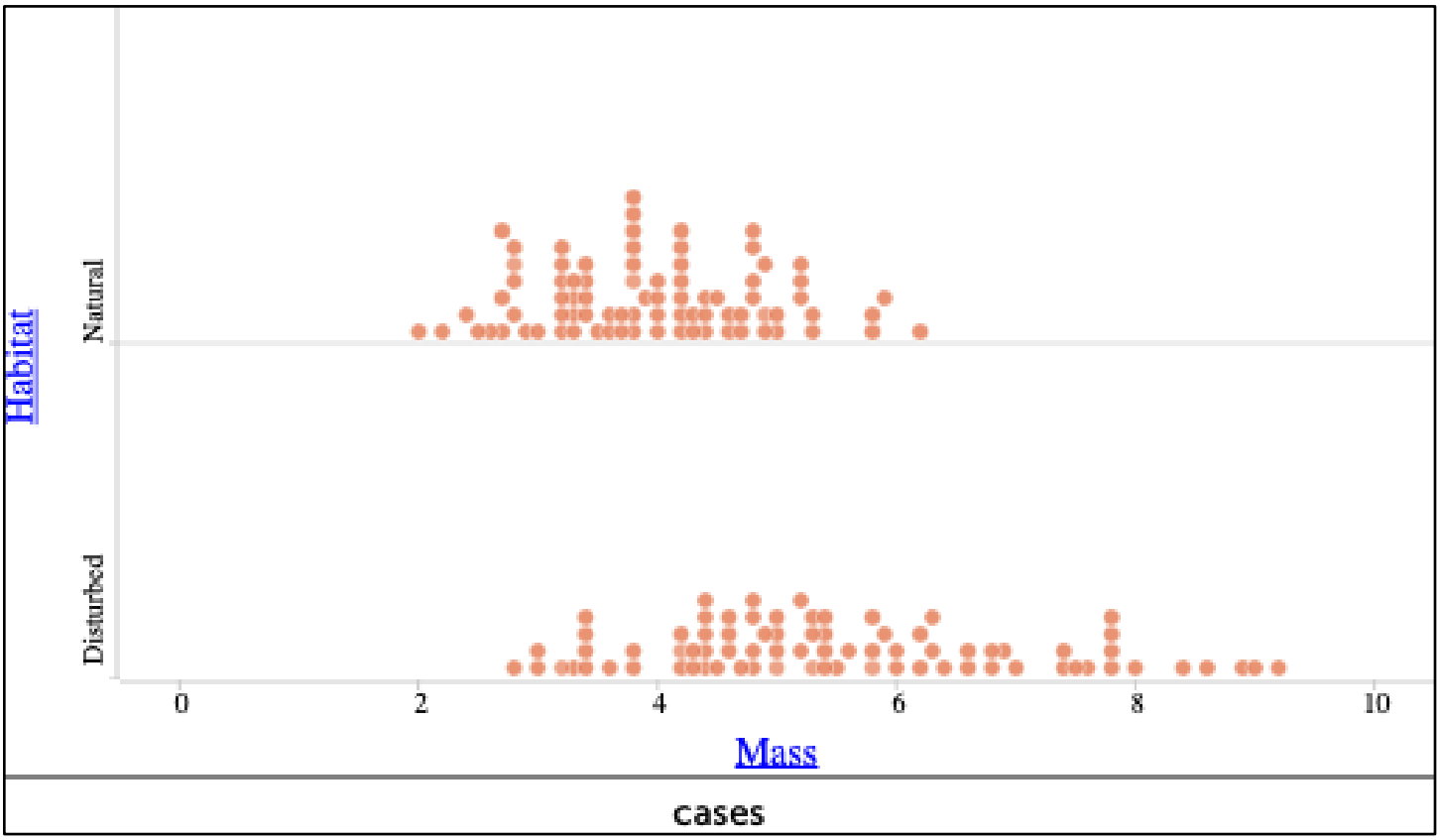
What is the misclassification rate for this new cutoff line?
41 = ***26%***
160
What does this misclassification rate tell us? Is there any improvement over the Trail Rule #1 rate?

[Type here]

27% of the lizards captured would be misclassified as natural if using the rule that the mass is less than 5 grams to classify them as from a nature habitat. This is a decrease of 8 percentage points from the previous rule.

Can we improve the misclassification rate by developing a classification rule using mass?
It seems very likely the misclassification rate can be improved upon.

Trial Rule #3: Create your own cutoff line in an effort to improve and optimize the misclassification rate.



Ask students to create at least one other cutoff line and calculate the misclassification rate for it. Compare student work as a class to determine the best cutoff line within those suggested using a document camera or pictures of the student work.

EVALUATE

Interpret

Use the best cutoff line with the smallest misclassification rate to create an interpretation of the analysis to answer the statistical investigative question and cite evidence from the analysis.

Statistical Investigative Question: **How can a lizard's mass be used to predict whether it came from a "Natural" or a "Disturbed" habitat?**
Interpretation: **A lizard's mass that is less than _____ will classify it as coming from a natural habitat with a misclassification rate of _____.**

[Type here]

Upon further reflection, it seems like it may be possible to use more data to be even more precise in classifying the lizards. So, there is a new challenge: **Is it possible to reasonably classify a lizard using two predictor variables instead of just one? How might we include head depth?**

EXPLAIN and ELABORATE - Part 2: Level C

Ask

The students are interested in exploring the following NEW statistical investigative question: **How can a lizard's mass and head depth be used to predict whether it came from a "Natural" or a "Disturbed" habitat?**

Collect

The students will continue to use the secondary data from the data set where a biologist captured (randomly) a number of individual lizards across two habitats on four islands in the Bahamas. The biologist took various measurements including mass, head depth, and hindlimb length.

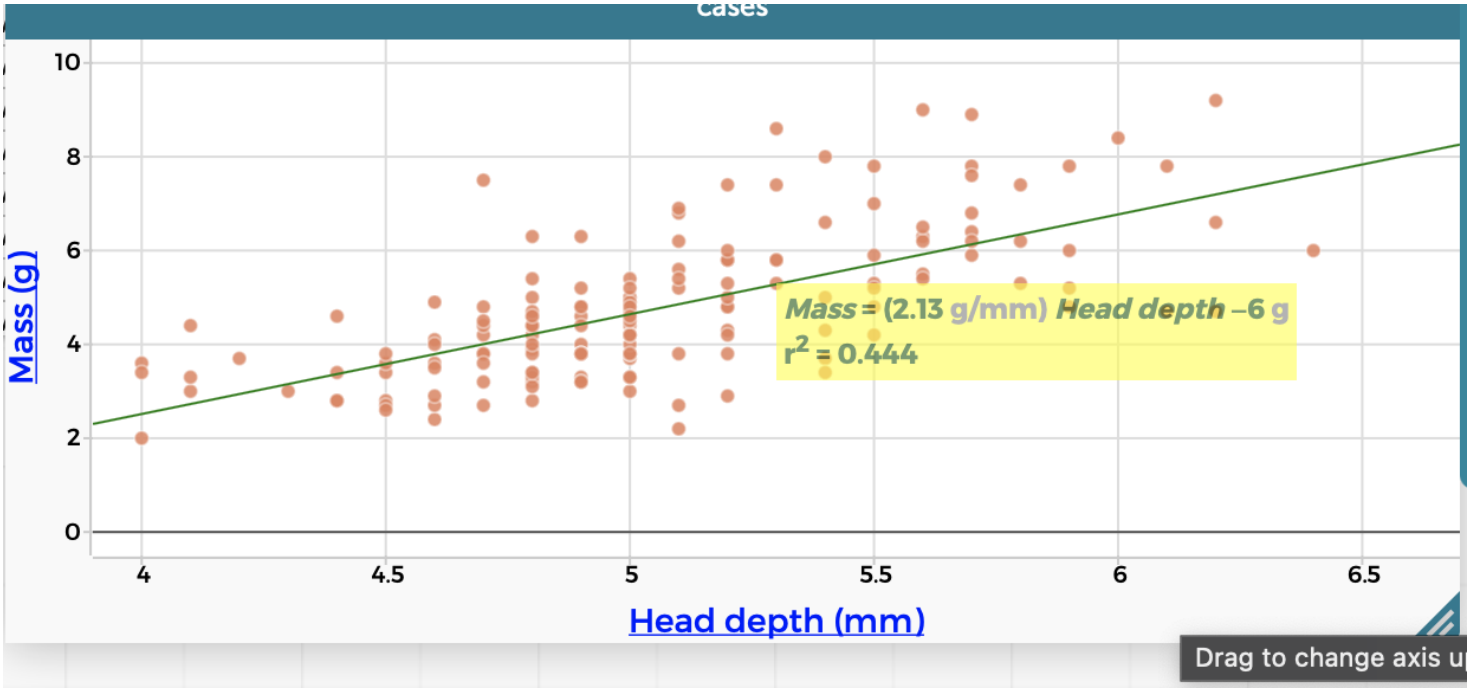
81 lizards in the study came from natural habitats.

79 lizards in the study came from disturbed habitats.

EXPLAIN and ELABORATE

Analyze

Let's explore first the bivariate relationship between head depth and mass.

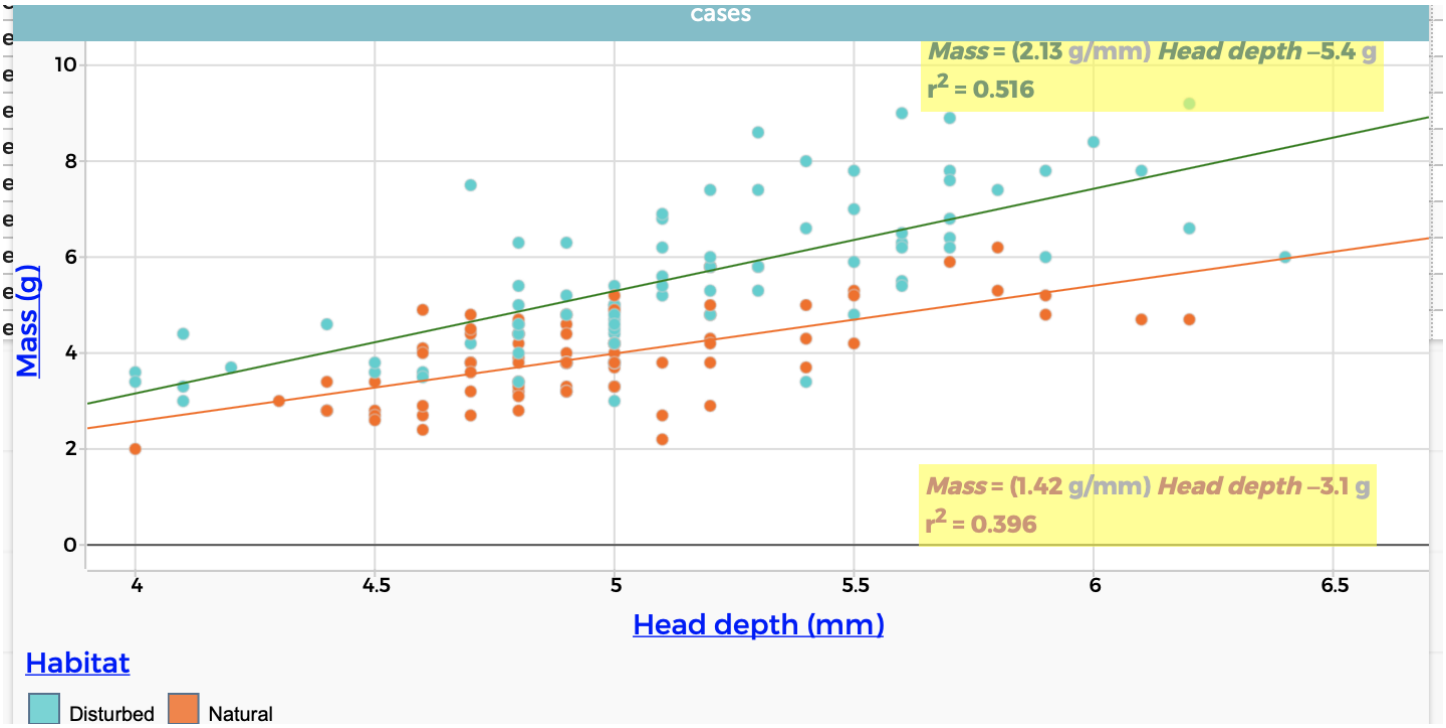


Ask students to answer the following interrogative questions:

- Is there a linear relationship between mass and head depth?
- Is the relationship positive or negative?
- What is the strength of the relationship?
- What is a possible fitted line to the data?

This graph looks at the bivariate relationship between head depth and mass. How might we incorporate the variable 'type of habitat' into this analysis? How might we fine tune our analysis to incorporate the type of habitat? Remember our NEW statistical investigative question: **How can a lizard's mass and head depth be used to predict whether it came from a "Natural" or a "Disturbed" habitat?** Use Think-Pair-Share as students work in partner pairs to make suggestions on how to incorporate type of habitat into the analysis.

We will utilize multivariate analysis, creating a new scatterplot of data that is grouped by habitat with the variables head depth and mass. We introduced the categorical variable "type of habitat" into the analysis by color coding the points on the scatterplot using technology. We observe the "natural" blue squares are clustering together and the "disturbed" orange squares are clustering together. What would happen if we fit two separate lines? Can we improve our prediction?



Using the open source applet from CODAP, two separate best fitting lines are found. It appears by having two separate fitted lines, we are able to better minimize the variability as compared to using just one fitted line.

Suppose a randomly chosen lizard has mass =3.6 g and head depth = 5.5 cm. Would you predict this lizard to be from a natural or disturbed habitat? How would you use the previous analysis to make a decision?

Ask students to work in partner pairs to explore this question and then discuss their responses with the class.

Sample responses:

- *We could plot the point on the scatterplot and make a decision either by which cluster the point falls closest to or which line it falls closest to.*
- *We could evaluate the two equations of the lines. $\hat{y}$(disturbed) = 5.05. $\hat{y}$(natural) = 4.575. Since the residual (error) is 0.45 for disturbed compared to 0.925 for natural, we would predict the lizard came from a disturbed habitat.*

Note: Residuals are the error term – measuring the variability around the fitted line by finding the difference in the actual y-value for a given x-variable and the *y*^-value. *y*^ (y hat) is the y-value predicted by the least squares regression line. If previous math courses do not cover residuals, then students may now utilize the smallest difference in the y^-value and the y-value for a given x-variable without calling it a residual. In other words, students may use the head depth value (x-value) in both least squares regression lines to find the predicted mass value of the lizard. The smallest difference in the predicted mass value and the actual mass value of 3.6 grams would suggest the type of habitat from the least squares regression line. This concept can then be named as a residual and added to the TIP Chart.

EVALUATE

Interpret

Using the two clusters of types of habitats with the variables of head depth and mass, answer the new statistical investigative question and cite evidence from the analysis.

NEW Statistical Investigative Question: **How can a lizard's mass and head depth be used to predict whether it came from a "Natural" or a "Disturbed" habitat?**

Interpretation: _____

_____

_____

_____

_____

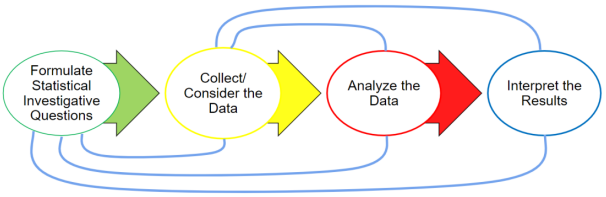
Students need to explain how they will use the clustered scatterplot to interpret and answer the new statistical investigative question.

Extension:

- Provide the Lizards data set to the students to explore using CODAP.
- Ask students in small groups to choose other variables of interest from this data set to create and answer new statistical investigative questions about classifying random lizards as from either a disturbed habitat or a natural habitat.
- Be sure students use and follow the Statistical Problem-Solving Process as they investigate.

Lizards Task Student Version

Levels B and C



Students in a biology science class have been studying about disturbed habitats (those that have human development) and natural habitats for lizards. They are interested in knowing if the mass of a lizard is useful in predicting the type of habitat a lizard lives.

What do you notice about the scenario described here?

_____

_____

_____

What do you wonder about the scenario described here?

_____

_____

_____

What statistical investigative question would you like to explore about the lizards?

_____

Ask

The students are interested in exploring the following statistical investigative question: **How can a lizard's mass be used to predict whether it came from a "Natural" or a "Disturbed" habitat?**

Interrogative Questions:

- Is it a specific type of lizard or any lizard?

- Is it for a specific geographic region?

- What is meant by mass?

Collect

The students use a data set where a biologist captured (randomly) a number of individual lizards across two habitats on four islands in the Bahamas. The biologist took various measurements including mass, head depth, and hindlimb length. The students are considering a *secondary data set* – data collected by another researcher in the Bahamas. If the students had collected the data, the data would classify as *primary data*.

	
81 lizards in the study came from natural habitats.	79 lizards in the study came from disturbed habitats.

Interrogative Questions:

- What type of lizards were used in the study?

- What are specific details about the Bahamas geography? How are the four islands similar/different?

- How were mass, head depth, and hindlimb length measured? What units were used?

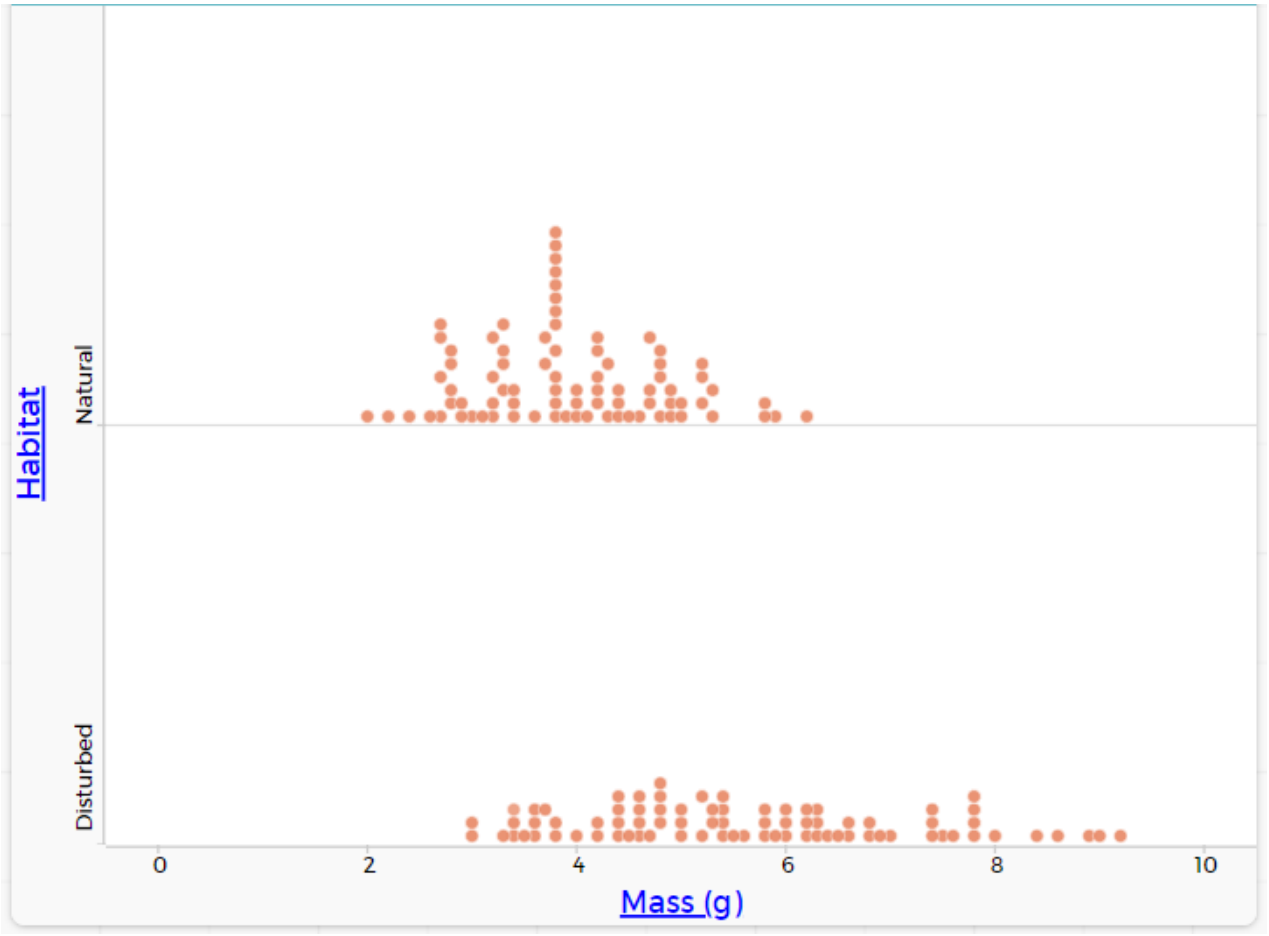
[Type here]

- What was the reason the lizards were randomly selected?

- Can this secondary data set reasonably answer our statistical investigative question?

Analyze

Here is a dotplot representation of the distributions using mass of the lizards and the habitats of the lizards collected in the biologists' study.



What do you notice about the dotplots?

_____

_____

_____

What do you wonder about the dotplots?

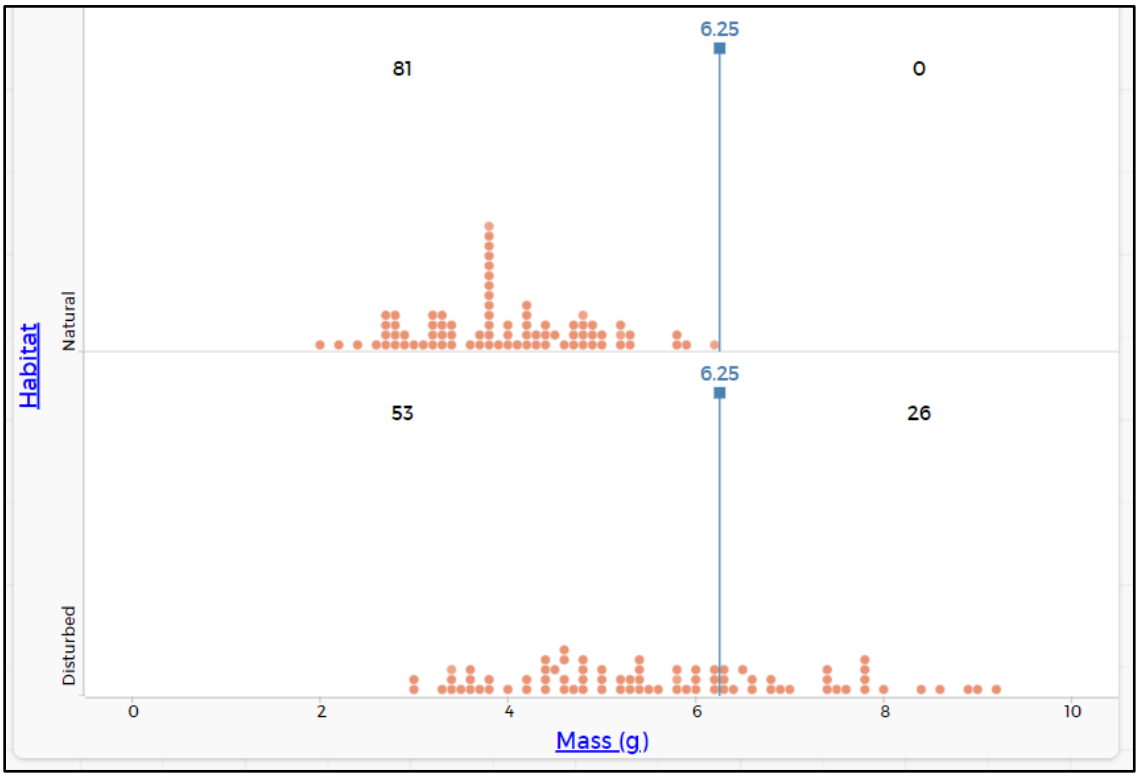
_____

_____

[Type here]

Where would you draw a vertical cutoff line on the dotplot representation to classify lizards as having come from a natural habitat?

Trial Rule #1: Classify all lizards with a mass less than 6.25g as natural. Classify all lizards with mass less than 6.25g as natural to the two distributions, since there is obvious separation between the mass of the lizards in the natural and disturbed habitats at 6.25 grams.



What do the numbers (n=81, n=0, etc.) on the dotplot representation mean?

But now we ask the analysis question, "How often do we misclassify the lizards based on using this classification rule?"

$$\text{Misclassification Rate} = \frac{\text{Number of Lizards Misclassified}}{\text{Total Number of Lizards}}$$

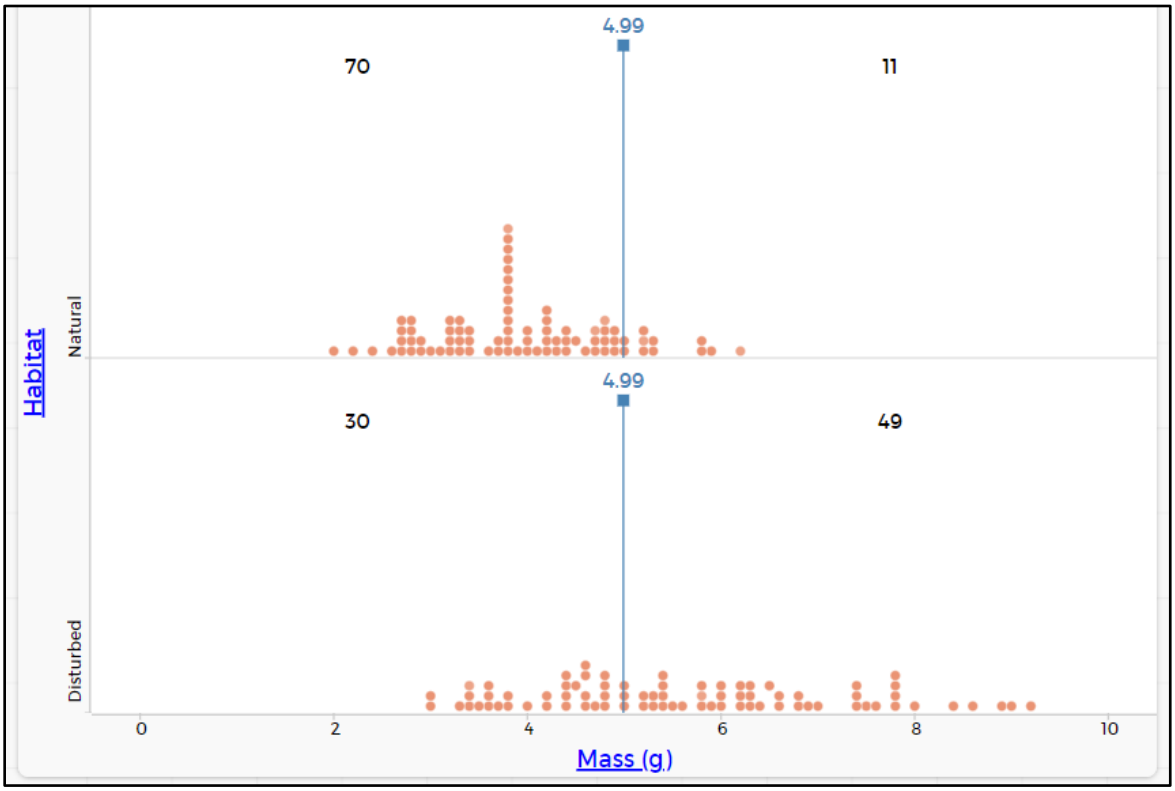
	Truly from natural habitat	Truly from disturbed habitat
Classified as "natural"	81	53
Classified as "disturbed"	0	26

What is the misclassification rate for our vertical line drawn at 6.25 grams?

[Type here]

What does this misclassification rate tell us?

Trial Rule #2: Classify all lizards with a mass less than 4.99 g as natural.



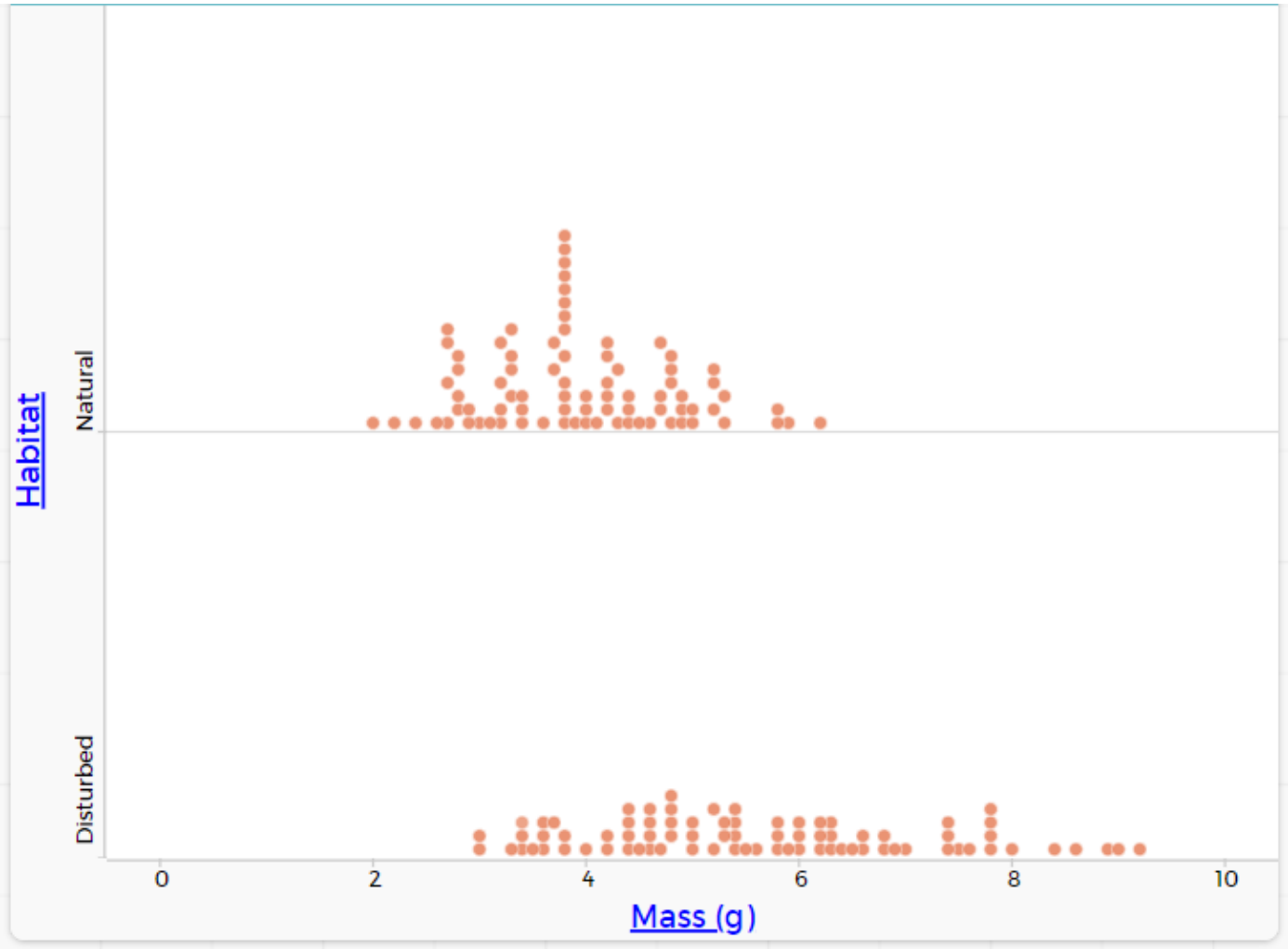
	Truly from natural habitat	Truly from disturbed habitat
Classified as "natural"		
Classified as "disturbed"		

What is the misclassification rate for this new cutoff line?

What does this misclassification rate tell us? Is there any improvement over the Trail Rule #1 rate?

Can we improve the misclassification rate by developing a classification rule using mass?

Trial Rule #3: Create your own cutoff line in an effort to improve and optimize the misclassification rate.



Vertical cutoff line:_____

	Truly from natural habitat	Truly from disturbed habitat
Classified as "natural"		
Classified as "disturbed"		

Misclassification Rate: _ _____

Interpret

Now compare your cutoff line with others in your class. Which is the best cutoff line for classifying lizards according to the mass of the lizard? How do you know? Cite evidence from your analysis.

Statistical Investigative Question: **How can a lizard's mass be used to predict whether it came from a "Natural" or a "Disturbed" habitat?**

Interpretation: _____

_____

Upon further reflection, it seems like it may be possible to use more data to allow for more precision in classifying the lizards. So, there is a new challenge: **Is it possible to reasonably classify a lizard using two predictor variables instead of just one? How might we include head depth?**

PART 2

Ask

The students are interested in exploring the following NEW statistical investigative question: **How can a lizard's mass and head depth be used to predict whether it came from a "Natural" or a "Disturbed" habitat?**

Collect

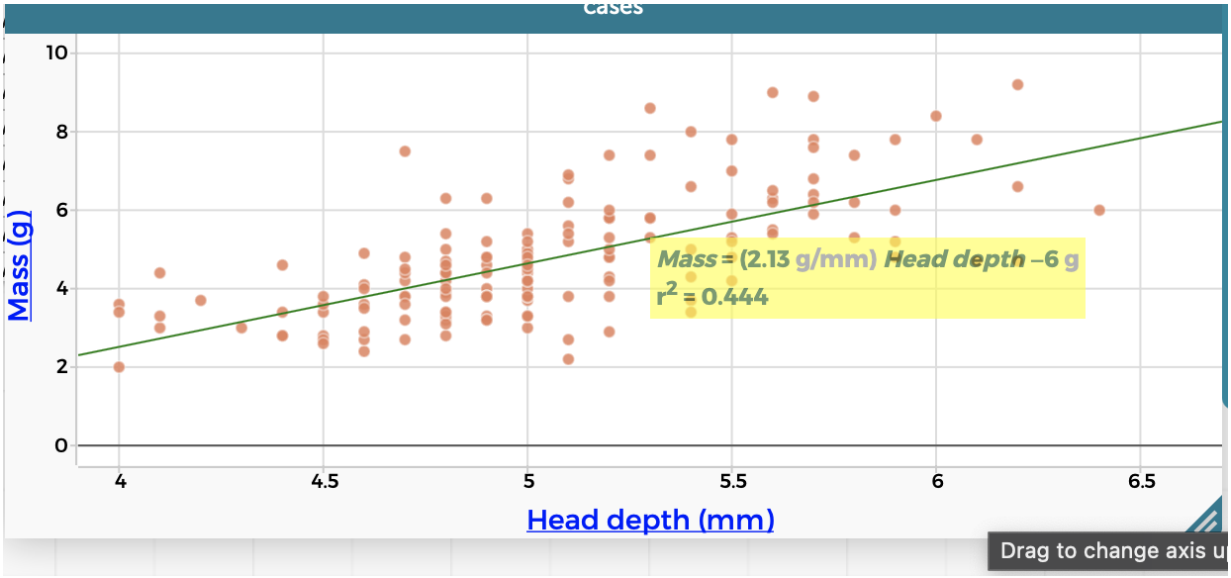
The students will continue to use the secondary data from the data set where a biologist captured (randomly) a number of individual lizards across two habitats on four islands in the Bahamas. The biologist took various measurements including mass, head depth, and hindlimb length.

	
81 lizards in the study came from natural habitats.	79 lizards in the study came from disturbed habitats.

APPLY

Analyze

Let's explore first the bivariate relationship between head depth and mass.



Answer the following interrogative questions:

- Is there a linear relationship between mass and head depth?

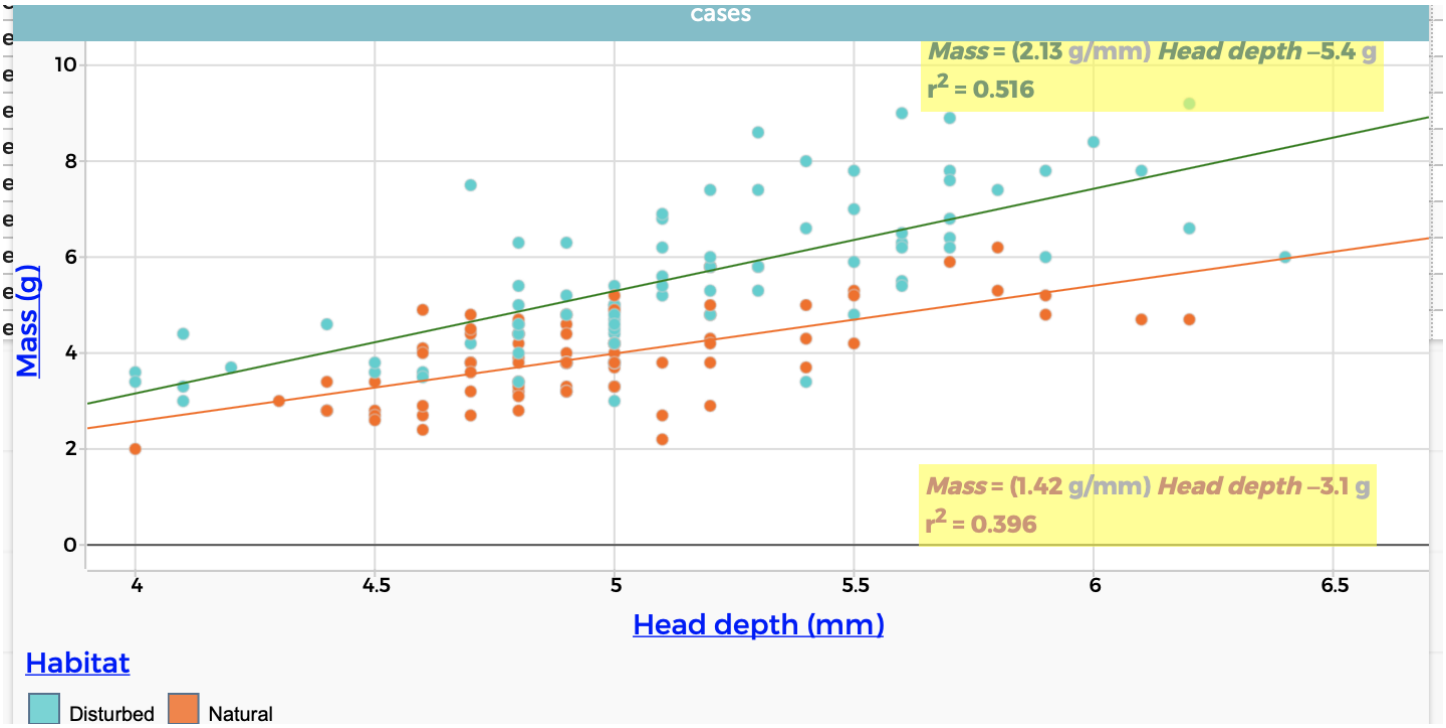
- Is the relationship positive or negative?

- What is the strength of the relationship?

- What is a possible fitted line to the data?

This graph looks at the bivariate relationship between head depth and mass. How might we incorporate the variable 'type of habitat' into this analysis? How might we fine tune our analysis to incorporate the type of habitat? Remember our NEW statistical investigative question: **How can a lizard's mass and head depth be used to predict whether it came from a "Natural" or a "Disturbed" habitat?**

Introducing the categorical variable "type of habitat" into the analysis by color coding the points on the scatterplot using technology, we observe the "natural" blue squares are clustering together and the "disturbed" orange squares are clustering together. What would happen if we fit two separate lines? Can we improve our prediction?



Suppose a randomly chosen lizard has mass =3.6 g and head depth = 5.5 cm. Would you predict this lizard to be from a natural or disturbed habitat? How would you use the previous analysis to make a decision?

REFLECT

Interpret

Using the two clusters of types of habitats with the variables of head depth and mass, answer the new statistical investigative question, citing evidence from your analysis.

NEW Statistical Investigative Question: **How can a lizard's mass and head depth be used to predict whether it came from a "Natural" or a "Disturbed" habitat?**

Interpretation: _____

_____

_____

_____

_____

_____