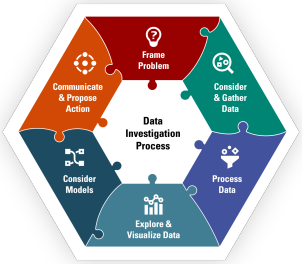
The Thrills of Roller Coasters: Using Data to Make Recommendations

Author(s):	Hollylynne Lee hollylynne@ncsu.edu Gemma Mojica gmmojica@ncsu.edu Emily Thrasher epthrash@ncsu.edu Zachary Vaskalis ztvaskal@ncsu.edu Bruce Graham bmgraha2@ncsu.edu
Affiliation and Location:	HI-RiSE: A Hub for Innovation and Research in Statistics Education Friday Institute for Educational Innovation, NC State University, Raleigh, NC This lesson plan and associated research was developed as part of the InSTEP project, funded by the National Science Foundation (DRL 1908760).
Abstract to Lesson Plan	This lesson builds on research of how real world data investigations can be used to engage students to learn key concepts in data science and statistics. The lesson uses a free online data visualization tool (CODAP) and a dataset of 635 roller coasters from US amusement parks in operation in Spring of 2020. Variables like top speed, material used, seating arrangement, length of track, and height quickly engage students in making sense of the data and using their real world and scientific understandings to make sense of trends and patterns in the data.
Learning Objectives	Students will be able to: ● Identify patterns and relationships they see in data ● Pose questions about relationships between different characteristics of roller coasters ● Process and Analyze data to answer their questions using statistical software ● Obtain, evaluate, and communicate information about their investigation and findings in written and oral form
Grade Levels	Content: Engineering, Math, Statistics and Physical Science grades 6-8, 9-12
Time	2-3 90-minute class periods
Resources Needed	A laptop or computer for every 2-3 students to share. Computer and projector for display Internet access (to use data and online tool linked below) Student Handout for Pre-Assignment Student Handout for Investigation (Roller Coaster Investigation) Handout about definitions of variables (Roller Coaster Data: variables and Definitions)
Apps/Websites Needed	In this lesson, students use the free online data tool CODAP [Common Online Data Analysis Platform] http://codap.concord.org The data includes a total of 635 roller coasters in operation in the U.S. in Spring 2020. For this lesson, on day 1, students start by exploring a few variables (characteristics) of roller coasters in a geographic region that includes their home state: South, NorthEast, MidWest, and West. Then on Day 2 of the lesson, they use the same CODAP file and *expand* the dataset to view all 635 cases and several additional variables (9 more variables). Document with coasters in the **South** visible: 14 states, 223 coasters, 12 variables Document with coasters in the **West** visible: 9 states, 122 coasters, 12 variables Document with coasters in the **Midwest** visible: 10 states, 134 coasters, 12 variables Document with coasters in the **Northeast** visible: 7 states, 156 coasters, 12 variables

Data Sources	Data primarily comes from: Roller Coaster Database. Missing data was acquired through other sources such as: Ultimate Roller Coaster, Coasterpedia, Coaster Grotto, and Wikipedia
Background Preparation for Teacher	This lesson organizes students' explorations into roller coasters using six key processes for data investigations: Frame the Problem, Consider and Gather Data, Process Data, Explore and Visualize Data, Consider Models, and Communicate and Propose Actions. This Data Investigation Process builds on frameworks from statistics education by including practices and processes from data science and other professionals that work with big data.  1. Read about this 6-phase Data Investigation Process http://cdn.instepwithdata.org/DataInvestigationProcess.pdf 2. When assisting students in a data investigation, it can be useful to help them develop key ways of thinking and dispositions that are helpful in developing expertise in conducting data investigations like a statistician or data scientist. This brief reading can guide you as a teacher in thinking through the data investigation process for you and your students. https://cdn.instepwithdata.org/ThinkingDataInvestigationProcess.pdf 3. If you make data investigations a central part of your teaching practices, it can be useful for students to see a reminder of this process and key considerations. There is a poster version of the Data Investigation Process that can be used in your classroom for this purpose. http://cdn.instepwithdata.org/DataInvestigationProcessPoster.pdf 4. In this lesson, students use CODAP, a free online dynamic statistics software tool. For a brief introductory video to learn how to use CODAP follow this link: https://youtu.be/aD5tLWld98w 5. Launching data-rich tasks involves hooking students into the context and ensuring they activate their real world understandings. You can watch a video of a teacher introducing this lesson by having students experience a Point of View video of a roller coaster ride. https://youtu.be/aXolxokHRxU 6. As you think about how to manage the different ways students may investigate roller coasters using multivariate data, it may be helpful to watch a brief video of how a teacher chooses students to share their coaster investigation and orchestrates a discussion using students' work: https://youtu.be/ETNF_542DvU

The Thrills of Roller Coasters: Using Data to Make Recommendations

Detailed Lesson Plan

Pre-Assignment (15-20 minutes)

The pre-assignment can help students to start to understand the phenomena of roller coasters and how coaster enthusiasts have created extensive websites for chronicling and archiving data about roller coasters. Students will have an opportunity to explore several websites about roller coasters. Specifically, each student will gather data about characteristics of a single roller coaster. They should choose a roller coaster that is somewhat familiar to them, either because of its geographic location, or their personal experiences.

Use the handout: **Roller Coaster Pre-Assignment**

Students should bring their completed handout to class.

Part 1 Lesson Plan (90 minutes)

Students should work in small groups (2-4). If pairs, they could share a single laptop, or each student could have their own. Everyone will need internet access.

Part 1 of the lesson is intentionally structured to assume students may be learning features of CODAP.

Framing the Problem (10 minutes)

Context: As a whole class, take a few minutes to engage your students in a conversation about roller coasters and stories of roller coaster enthusiasts. The following gives some stories to share:

Several stories have been featured in the media about roller coaster enthusiasts Drs. Cheryl Schreiber Lewison and Martin Lewison, a physician and college professor from New York. These coaster enthusiasts travel the globe to ride over 100 roller coasters a year! They have ridden over 2,000 roller coasters around the world, and have shared their experiences and tips about their favorite coasters. There are many websites that chronicle the experiences of roller coaster enthusiasts and record aspects of these experiences and characteristics about coasters. There is a database of thousands of roller coasters from around the world, Roller Coaster DataBase, featuring information about *speed*, *height*, *track length* as well as other features.

Purpose: As a whole class, discuss the purpose of the next two days of lessons with your class:

Because of the enthusiasm for roller coasters, our class has been asked to write blog posts sharing insights about roller coasters. Our job is to help people understand the characteristics of different kinds of roller coasters in our region of the United States, and perhaps how characteristics of coasters in our region may compare with those in other regions. (Note: See this map for how the United States is separated into four generally recognized regions: West, Midwest, South, Northeast).

Understanding the Phenomena: Students need to understand roller coasters in order to be able to help with the blog. As a whole class, dig into ideas about the roller coaster experience and different variables of roller coasters. The following are some resources to help guide this discussion.

Not everyone has ridden a roller coaster. Below are two point of view (POV) videos that capture the experiences of riding two different roller coasters. The first is a wooden coaster called the Jack Rabbit at the Seabreeze Amusement Park in New York. Even though it is one of the oldest US roller coasters that is still in operation, it was the fastest at the time it was built. The other roller coaster is Kingda Ka, a steel coaster located at Six Flags Great America in New Jersey, which was the tallest and fastest coaster in the world when it opened. Before you watch both videos, ask students to consider the following question: *What physical aspects of roller coasters make them exciting or scary?*

Jack Rabbit video (1:55 minutes) Kingda Ka video (0:28 minutes)

Lead a discussion with your class about **What physical aspects of the roller coasters that made them exciting or scary?** Some variables that students may bring up are: *speed*, *maximum height*, *drop*, *inversions*, *g-force* experienced, *length* of the track, *duration* of the ride, *age* of the coaster and *type* of track. The goal here is to activate your students' thoughts about different variables of roller coasters to prepare them to understand the dataset they will work with later and get them curious about how these variables relate to the roller coaster experience. Here are possible follow up questions to help guide the whole group discussion:

1. Do you get a sense of some key differences in riding the coaster with the wooden track versus steel track? What differences did you notice?
2. After considering the question above and watching the POV videos, what do you wonder about related to coasters?
3. Is there anything you are curious about after watching the videos or from experiences you have with roller coasters?

Considering and Gathering Data (15 minutes)

Using the data that students gathered in the pre-assignment, ask several students to share the roller coaster they found and some of its characteristics with their partners. Then have a whole group discussion about some of the following ideas:

- Who found a roller coaster that seems to be pretty fast? Did anyone find a roller coaster that was faster than that one? How do we know what should be considered fast for a roller coaster?
- Do any students have missing data about their roller coaster? Did you have trouble finding some of the characteristics of a roller coaster? Did you have to go to more than one website to find information?
- Did anyone find conflicting measurements on different websites about their roller coaster? Why might there be different measurements on different websites?

Have students imagine collecting all of their roller coasters and the measurements they found into a collection. This would be a dataset with “X” roller coasters (X=number of coasters found by students in the class). For each case in the dataset, students found information about 12 different characteristics of a roller coaster

(Name, Amusement Park, etc) that we are going to consider to be attributes (or variables). See the table in the Pre-Assignment for the list of all attribute names. Collecting their own roller coaster data helps students understand where data comes from and the structure of a dataset.

On the student handout (**The Thrills of Roller Coasters: Using Data to Make Recommendations Part 1**) have students discuss and respond to #1.

1. What physical variables make roller coasters exciting or scary? What do you wonder about roller coaster features?

Identify the data source and collection methods: The larger dataset for this lesson consists of US roller coasters that were in operation in Spring of 2020. There are 635 coasters with 21 categorical and quantitative variables. Take a few minutes to help students understand how the dataset they will be using was collected. At the time data was collected, all operating US coasters, according to the Roller Coaster Database, were identified. The Roller Coaster Database claims to contain all operational roller coasters. Statisticians, data scientists and other professionals that work with data often spend much of their time Processing the Data (e.g., organizing, structuring, cleaning, transforming). After data was obtained for all 635 roller coasters in operation, there was a lot of missing data values for several of the attributes. Thus, additional data sources were used to gather as much missing data as possible: Ultimate Roller Coaster, Coasterpedia, Coaster Grotto, Wikipedia. Like the Roller Coaster Database, these sources all contain lots of current information about roller coasters. These sources provide current and (mostly) accurate information and are considered to be trustworthy and generally reliable sources, which should always be questioned when gathering or considering data that has already been collected.

To begin our investigation, we are going to focus on a subset of the data that only includes roller coasters in a specific region of the United States. We suggest that your class start with the data file for the region containing your home state. There are some states (e.g., Montana, Mississippi) that had no roller coasters in operation.

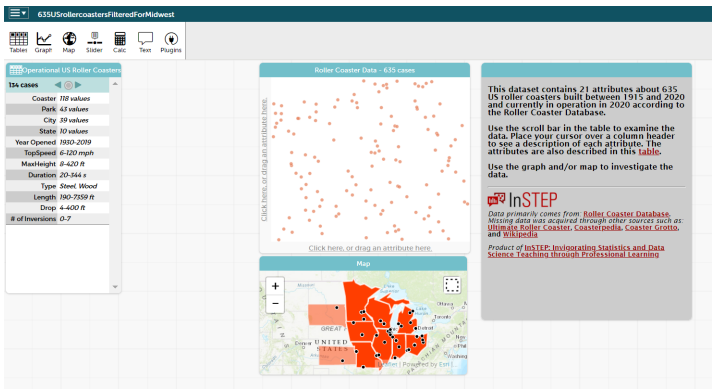
Document with coasters in the **South** visible: 14 states, 223 coasters, 12 variables

Document with coasters in the **West** visible: 9 states, 122 coasters, 12 variables

Document with coasters in the **Midwest** visible: 10 states, 134 coasters, 12 variables

Document with coasters in the **Northeast** visible: 7 states, 156 coasters, 12 variables

Once students open a data file in CODAP, they will see four windows that have been prearranged for them. All data is contained in the cards at the far left of the screen. There is also a Map showing the location of the roller coasters in th specific region, a graph window that has scattered case icons randomly places (each dot represents a roller coaster), and gray box that contains some basic information about the data in this file.



In CODAP, you can view a data collection in either a Table or a Case Card view. The roller coasters are initially seen in a *case card view*. Imagine this as a stack of index cards where each card contains information about different variables for a case. Connect this back to the data sheets they used to organize their data in the pre-assignment. Imagine stacking these sheets on top of one another.

In the case card view you can use the arrows to browse each case of the dataset. You can toggle to *case table view* by clicking the icon in the upper left corner of the case card view to see the cases together in a table, where each *row* is a case of a roller coaster, and *columns* contain information about each variable.

Give students a few minutes to complete 2-6 on the handout.

Posing an Initial Investigative Question (2 minutes)

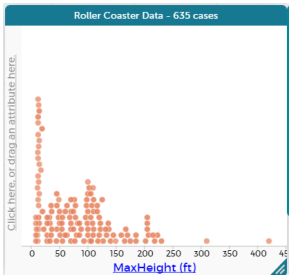
To get everyone started, have them consider an initial question:

How tall do roller coasters in your region tend to be?

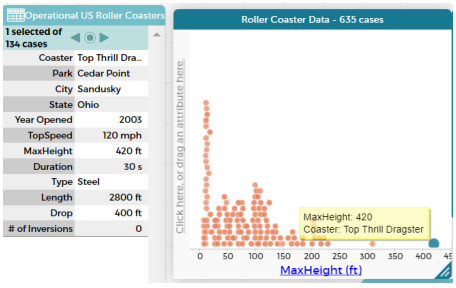
Ask what variable, or characteristic of a roller coaster, could be used to answer this question (Max Height).

Exploring the Data (5 minutes)

Demonstrate to students how to create their first graph by following directions on the student handout. Below is an example of a dot plot showing the distribution of the MaxHeight for roller coasters in the Midwest.



Students are often first attracted to examining extreme cases. Questions 6-7 help them explore this through using the multiple linking feature of CODAP. When they hover over a dot in the graph a tip box will appear with some information about that case. If they click on a dot in a graph, that case will be highlighted in the data collection (in case card or table view).



Have students complete #6-7 on the student handout to find out about some of the extremely tall or small coasters. (Hint: if data is clumped together on the number line, it may be difficult to see which case is the smallest. You can change the size of the graph window by resizing the window through draggin to expand the scale, or even drag the values on the axes to adjust the scale within the graph window.)

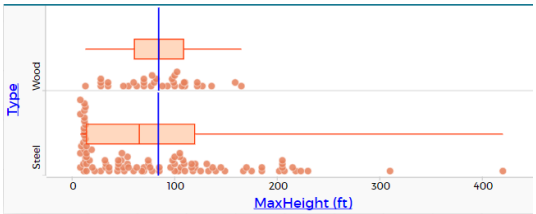
Considering Models (10 minutes)

To help describe the distribution of maximum heights of the roller coasters, students can explore different models that can be used. A dot plot, box plot, binned plot, and histogram are a few of the types of graphs useful for displaying a distribution of a quantitative variable such as max height. You can help students create

different graphs by exploring what features are available in the ruler and configuration menus on the right side of the graph menu (click in the graph window to activate this side menu). There are also tools in the ruler menu that can help students overlay moveable lines, create shaded regions, and add vertical lines indicating the location of the values for typical measures of center such as mean and mode.

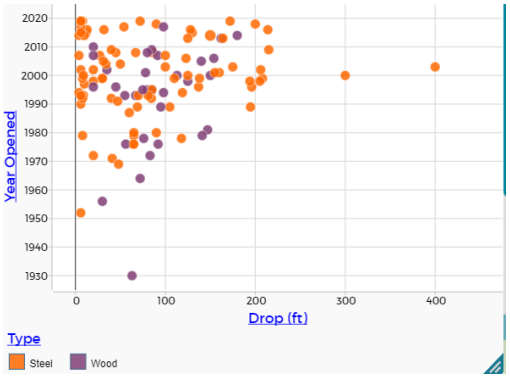
Have students complete # 8-9 on the student handout. You may need to assist them in learning to use some of the tools in the graph menu to accomplish each task.

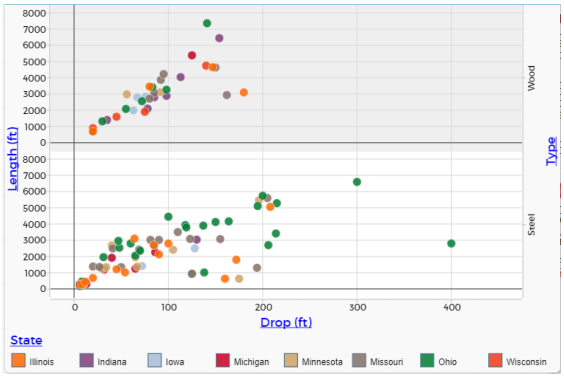
In question 9, students are asked to consider if there is a difference in max heights when you consider the type of material the coaster is made of. Here is an example of a graph of max heights of roller coasters in the midwest, separated by type. Although the have almost the same mean, the variability in steel is much greater, as seen with the boxplot.



Exploring the Data and Considering Models (15 minutes)

The next section of the handout (#10) encourages students to engage in a more open exploration of other variables in the dataset where students may have started wondering about or are curious to dig in further. Give students time to work in their small groups and monitor their progress by visiting each group. You may need to pose some questions to further their investigation or do some targeted instruction in ways they can visualize multiple variables in a graph. For example, be sure students know they can drag a variable name into the center of a graph and that this action will recolor the case icons according to the scale (numeric) or category (categorical) of that attribute. For example the graph on the left uses 3 variables, while the one on the right uses four variables (the fourth variable “Type” is added on the right side of the graph window).



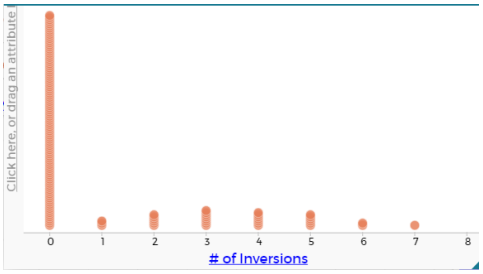


Communicating (10 minutes)

Select a few students who could share their graphs with the whole class and discuss what they found. This can be helpful in having students learn to communicate their findings about what they found to be interesting and for other students to learn various ways to use CODAP to make different graphical displays.

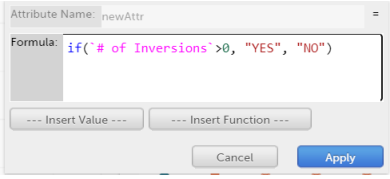
Creating a Variable (10 minutes)

[note: depending on the goals of your lesson and grade level of students, you may choose to skip this part]

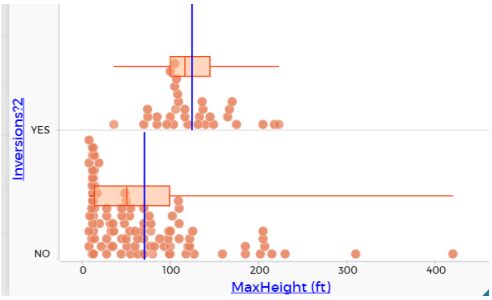


Some students may have used the variable " # of Inversions" in their prior exploration. The purpose of this part of the lesson is to give students an opportunity to experience how to transform data by using values of one variable to create a new variable. The variable # of Inversions is considered by default a numeric variable since it has values of 0, 1, 2, 3, ... A quick graph of the distribution of the # of inversions for coasters in the midwest shows that the vast majority of the coasters have 0 inversions. Meaning a rider never gets inverted. For a variable like this, it may be useful to compare coasters in two different groups: you never get inverted, and you do get inverted at least once.

In #11-13 on the student handout, students can explore how to add a new variable to the dataset by adding a new attribute in the table with the + symbol and then using an "if-then" formula that uses a logic statement to look at the value of the variable # of Inversions and then return either a YES or a NO based on whether it is greater than 0 (YES) or not (NO). The formula should be colored as in this image to indicate it was entered correctly. The proper syntax is important. You can use the HELP feature in the upper right of the CODAP menu if you need to search for help in writing a formula.



Graphing Max Height and separating by the new variable Inversions? should illustrate how coasters that invert riders (YES) tend to be taller but have a much smaller variability than those coasters where riders are not inverted.



Communicating Findings (13 minutes)

Have students spend time considering #14 on the handout. This is opportunity for them to share findings, and begin to write ideas they will use for their blog post about the roller coasters in their region of the U.S. This could be given as homework in preparation for the next day.

Part 2 Lesson Plan (90 minutes)

Students should work in small groups (2-4). It is recommended that they work in the same pairs/small groups as previously.

Part 2 of the lesson is intentionally student centered with facilitation by the teacher and provides more ways for students to investigate questions that interest them so that they can write a blog post about characteristics of roller coasters that they feel others would like to learn about. Have students use the handout **The Thrills of Roller Coasters: Using Data to Make Recommendations Part 2**

Revisiting Framing the Problem (5 minutes)

Begin the lesson by reminding students of the problem that will be guiding their investigations: Our class has been asked to write blog posts to share recommendations about roller coasters. Our job is to help people understand the characteristics of different kinds of roller coasters in our region of the United States, and perhaps how characteristics of coasters in our region may compare with those in other regions.

Ask students to share what they learned about characteristics of roller coasters from the previous part.

Considering Data/Processing Data (5 minutes)

Using the CODAP document of your region from the previous part, demonstrate how to view all 635 US roller coasters and attributes that are included in the dataset. The dataset consists of 635 coasters that were operational in the Spring of 2020, which included 21 categorical and quantitative variables.

Open the your CODAP document from that you previously worked with:

CODAP Document with coasters in the **South** visible: 14 states, 223 coasters, 12 variables

CODAP Document with coasters in the **West** visible: 9 states, 122 coasters, 12 variables

CODAP Document with coasters in the **Midwest** visible: 10 states, 134 coasters, 12 variables

CODAP Document with coasters in the **Northeast** visible: 7 states, 156 coasters, 12 variables

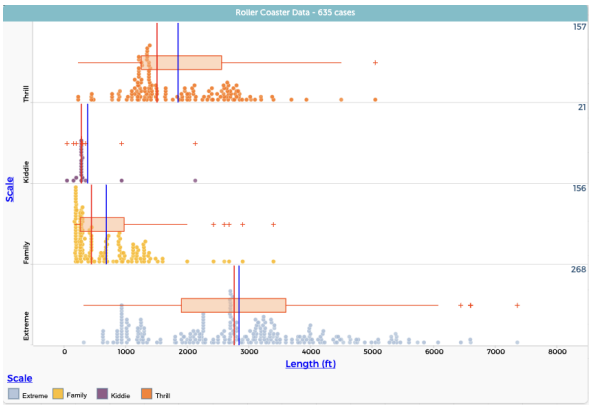
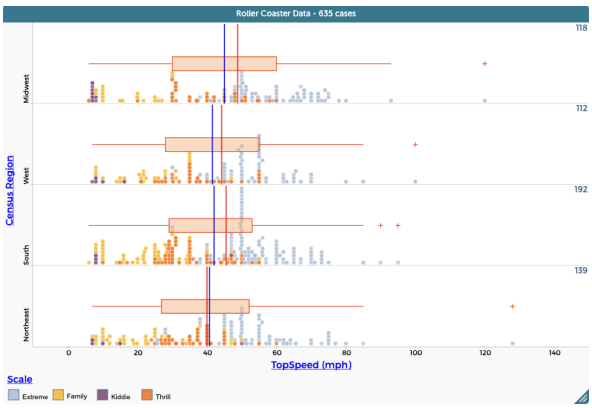
To view the dataset of 635 roller coasters, in the Table view, in the Right hand tools, Under the  menu, choose to “Restore Set Aside Cases” and to “Show Hidden variables”.

Investigating 635 Roller Coasters (80 minutes)

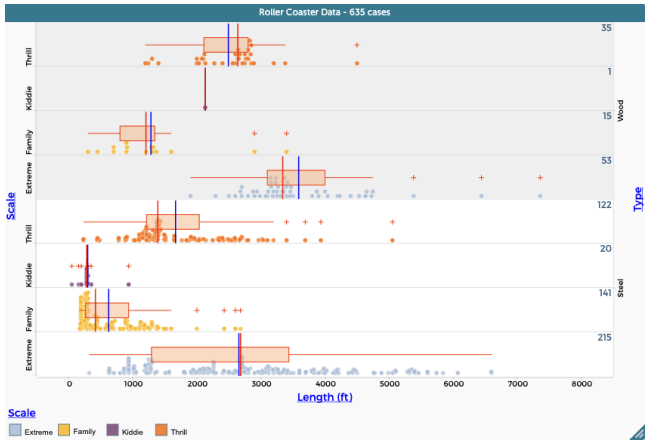
(10 minutes)
Ask students to examine the dataset in a Table view or Card view (#1). In a whole class discussion, ask students to share their noticings. You may want to highlight the following ideas or ask probing questions to get students to consider the following ideas. Students may notice:

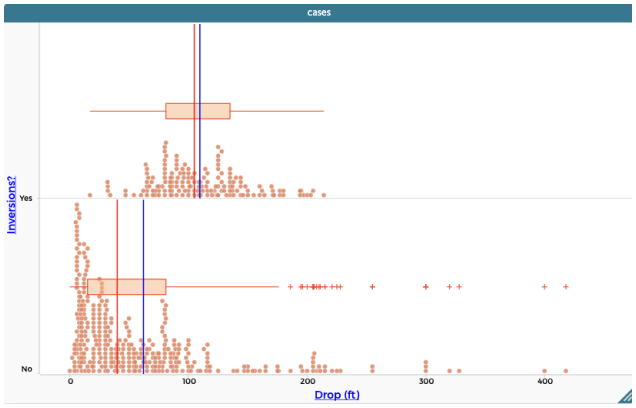
- there are more cases and more variables.
- there are both categorical and quantitative variables.
- what the variables measure and the unit of measurement (if specified).
- there is missing data for specific variables.

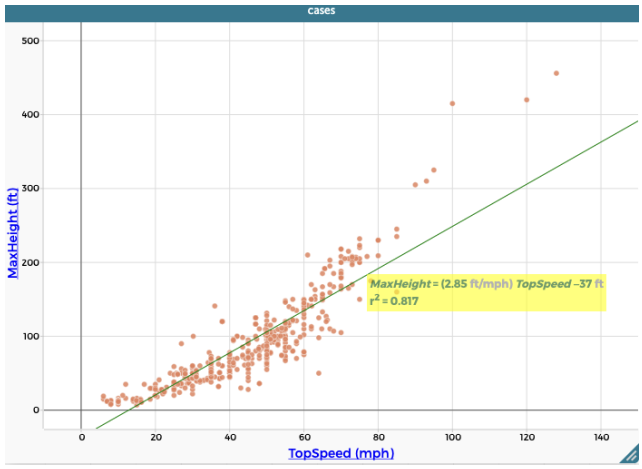
(15 minutes)
Ask students to choose one of the questions from (#2) to investigate in pairs/small groups. As students work in pairs/small groups, encourage students to create graphs and statistical measures to support any claims they make. As you monitor student work, select a few that highlight important ideas to share as a whole group. Ask a few pairs/small groups to share in the whole class discussion that highlights these ideas. Some investigative questions may lead students to clear findings and recommendations, but other questions may have less clear findings. See examples below.

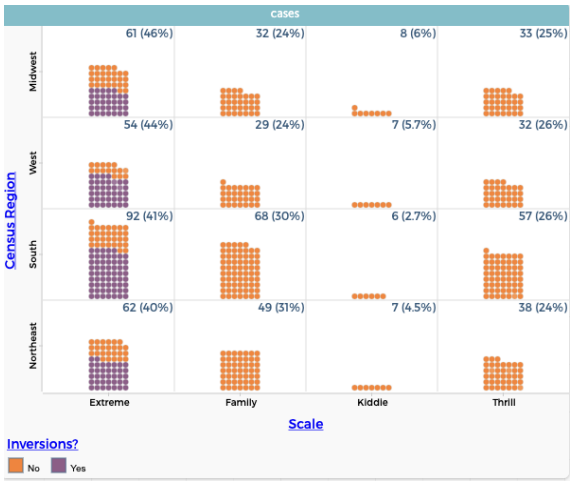
Possible student created graph for How does **Length** compare across roller coaster **Scale**?	
Possible student created graph for What **region** of the US tends to have the **fastest** roller coasters?	

(55 minutes)
Ask students to investigate questions (see #3 on student handout) related to characteristics of US roller coasters to help them make recommendations about characteristics of different kinds of coasters in your region and perhaps how characteristics of coasters in your region compare with those in other regions. As students work in pairs/small groups, encourage students to create graphs and statistical measures to support any claims they make related to their recommendations. Monitor their progress by visiting each pair/small group. Pose questions to further their investigation, as needed. The following include possible student graphs tat may be generated.









Communicating and Proposing Actions

Provide time for students to write and post their blog responses. Additionally, you can ask students to read and possibly respond to each other's blog posts. Most likely students may need extended time either inside or outside of class time to finish their blog post.

Roller Coaster Pre-Assignment

We are going to examine some data about roller coasters to consider some questions related to features of roller coasters and a rider's experience. To help get us started, each student needs to find *as much* of the following information about **ONE roller coaster**. Find data about a roller coaster at an amusement park that you have visited or that you may have heard about (i.e., there is something about the roller coaster or amusement park that is familiar to you).

Websites to find data about roller coasters include:

http://rcdb.com
https://www.ultimaterollercoaster.com/
https://coasterpedia.net/
http://www.coastergrotto.co.uk/

Bring your data to class. You can either print out the next page, or submit an electronic copy.

Coaster Name	
Amusement Park	
City	
State	
Year Opened	
Top Speed (mph)	
Maximum Height (ft)	
Ride Duration (seconds)	
Type of Material Used in Majority of Construction (Wood or Steel)	
Track Length (ft)	
Drop length (ft)	
Number of Inversions	
Is the roller coaster currently in Operation? (Yes or No)	
Sources used for Data (URLs)	
URL for a Picture or Video of this Coaster	
Is there anything about *this coaster* that you are curious or wonder about?	

The Thrills of Roller Coasters: Using Data to Make Recommendations
Part 1

Framing the Problem

Purpose
Our class has been asked to write blog posts to share recommendations about roller coasters. Our job is to help people understand the characteristics of different kinds of roller coasters in our region of the United States and perhaps how characteristics of coasters in our region may compare with those in other regions.

From watching the videos and the class discussion:

1. What physical variables make roller coasters exciting or scary? What do you wonder about roller coaster features?

Identifying and Considering the Data

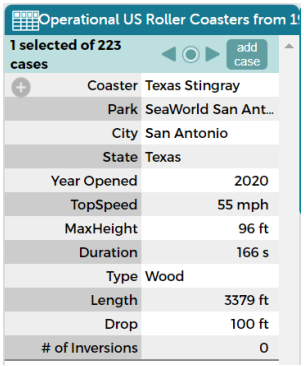
To begin our investigation, we are going to focus on a subset of the data that only includes roller coasters in a specific region of the U.S. Choose the data file for the region containing your home state. (Note: See this map for how the U.S. is separated into four generally recognized regions: South, West, Midwest, Northeast).

CODAP Document with coasters in the **South** visible: 14 states, 223 coasters, 12 variables

CODAP Document with coasters in the **West** visible: 9 states, 122 coasters, 12 variables

CODAP Document with coasters in the **Midwest** visible: 10 states, 134 coasters, 12 variables

CODAP Document with coasters in the **Northeast** visible: 7 states, 156 coasters, 12 variables



You can view a data collection in either a Table or a Card view. The roller coasters are initially seen in a *case card view*. Imagine this as a stack of cards where each card contains information about different variables for a case (the card shown is for the case of Texas Stingray at SeaWorld in San Antonio). In this view you can use the arrows to browse each case of the dataset. You can toggle to *case table view* by clicking the icon in the upper left corner of the case card view to see the cases together in a Table, where each row is a case of a roller coaster, and columns contain information about each variable.

2. Take a few minutes and play around with the dataset. Be sure to look at the data in both the Card view and the Table view. What do you notice and wonder about?

3. Click through the cards in the CODAP document, is the roller coaster you found in the pre-assignment in this dataset? What about your partner's roller coaster?

4. Comparing the Card view and Table view, which do you like better? Why?

5. Examine the CODAP file, where does this data come from? Do you consider this data trustworthy, why or why not?

Posing an Initial Investigative Question

How tall do roller coasters in your region tend to be?

Exploring the Data

In the shelf, click on the Graph icon 

to display a graph window. Each dot in the graph window represents a case of a roller coaster. To investigate how tall older roller coasters tend to be, let's graph the **MaxHeight** (drag label **MaxHeight** from either the case card or table to the x-axis in graph). Click on a dot and notice the table or case card. Explore the coasters by clicking on dots or selecting rows in the table or flipping through the case card view.

Find some information about the height of extreme coasters:

6. Name of coaster with the smallest maximum height: _____

   Smallest Maximum height: _____

   What else can you find in the data table about the coaster with the smallest height?

7. Name of coaster with the tallest maximum height: _____

   Largest Maximum height: _____

   What else can you find in the data table about the coaster with the tallest maximum height?

Considering Models

Now, let's describe how tall the roller coasters *tend* to be in your region. Describing tendencies for all roller coasters in a region involves considering different models that can be helpful in describing a distribution.

8. Add two moveable lines to form a shaded region and move it so it covers a subgroup of the coasters you consider to be typical in maximum height. (Use the ruler menu in the graph window to add moveable values.)

   a. What range of data does your shaded area cover?

   b. How many coasters are in the shaded region (look under the ruler menu)?

   c. What percent of data are in your shaded region (look under the ruler menu)?

Before moving on, remove the shaded region by going to the ruler menu and removing the movable values.

9. Explore adding a mean, median, and box plot to the graph (look under the ruler menu).
   a. Describe the typical maximum height for the roller coasters in your region.
   b. Do these coasters all have similar heights?
   c. Does the max height different for coasters that are made of different materials (Type: Wood or Steel)? (Tip: You can add Type to the horizontal axis)
   d. Based on this sample of data, what do you anticipate the typical maximum height of *all* roller coasters in the U.S.

Exploring the Data

10. Explore other variables that you wonder about related to roller coaster features to describe their tendencies. Provide evidence to support tendencies you find using graphs and statistical measures based on tools you learned in questions 8 and 9. Save your graphs and statistical measures and think about how you could use this to give recommendations to others about what they might expect about roller coasters in your region of the U.S.

Creating a Variable

A useful tool in an investigation is to create new variables based on existing ones. This is often a skill used when processing data. Notice that the variable **# of Inversions** is numeric (e.g., 0, 1, 2, ...). Sometimes it is helpful to change the way we use variables. Below you will find directions for creating a new variable.

11. If you did not explore the variable **# of Inversions** before, take a moment and graph this variable. What tendencies do you notice?

Instead of considering the number of times a rider is inverted (with the variable **# of Inversions**), it can be simpler to consider only two categories: if the riders ever get inverted or not. Create a variable with two categories indicating YES (riders are inverted), or NO (riders are never inverted).

a. Click on the + symbol in the upper right hand in the Table.
b. Name the new variable **Inversions?**.
c. Click the variable name at the top of the column, and select “edit formula”. Use the following formula: if(`# of Inversions`>0, "Yes", "No").
d. Click apply.

12. Graph your new variable **Inversions?.** What does this tell you?

13. Use the new variable to reconsider tendencies about maximum height. Create a graph of **MaxHeight**. Drag and drop **Inversions?** to the middle of your **MaxHeight** graph. What do you notice?

Communicating Findings

14. What have you learned so far about characteristics of roller coasters in your region? What might you want to tell others about recommendations for roller coaster experiences in your region?

The Thrills of Roller Coasters: Using Data to Make Recommendations
Part 2

Revisiting Framing the Problem

Our class has been asked to write blog posts to share recommendations about roller coasters. Our job is to help people understand the characteristics of different kinds of roller coasters in our region of the United States, and perhaps how characteristics of coasters in our region may compare with those in other regions.

Considering Data & Processing Data

Instead of examining roller coasters from a specific region, a dataset of all US coasters that were operational in Spring of 2020 will provide data to inform the blog post. The dataset consists of 635 coasters with 21 categorical and quantitative variables.

Start by opening the CODAP documents below that you worked previously:

CODAP Document with coasters in the **South** visible: 14 states, 223 coasters, 12 variables

CODAP Document with coasters in the **West** visible: 9 states, 122 coasters, 12 variables

CODAP Document with coasters in the **Midwest** visible: 10 states, 134 coasters, 12 variables

CODAP Document with coasters in the **Northeast** visible: 7 states, 156 coasters, 12 variables

To view the dataset of 635 roller coasters, in the Table view, in the Right hand tools, Under the menu, choose to “Restore Set Aside Cases” and to “Show Hidden variables”.

Investigating 635 Roller Coasters

1. Examine the dataset in a Table view or Card view. What do you notice about the cases and variables?

2. Choose one of the two questions below to investigate. Provide evidence using graphs and statistical measures to support any claims you make.
    a. How does **Length** compare across roller coaster **Scale**?
    b. What region of the US tends to have the fastest roller coasters?

3. Investigate other questions related to characteristics of US roller coasters to help you make recommendations about characteristics of different kinds of coasters in your region and perhaps how characteristics of coasters in your region may compare with those in other regions. Use your findings to write your blog post.

Roller Coaster Data: Variables and Definitions

Variable Name	Description
Coaster	Name of the roller coaster
Census Region	Identifies region of US in which state is located
Park	Name of the park where the roller coaster is located
City	City where the roller coaster is located
State	State where the roller coaster is located
Year Opened	Year when roller coaster opened
Age Group	Age intervals of when the coaster was built Age Groups: 1: Older (Built between 1900-1979) 2: Recent (1980-1999) 3: Newest (2000-current)
Top Speed (mph)	Maximum speed of roller coaster
Max Height (ft)	Highest point of roller coaster
Scale	Type of experience expected on the roller coaster (Kiddie, Family, Thrill, and Extreme)
Type	Material of track (Steel or Wooden)
Design	How a passenger is positioned in the roller coaster. Design Types:

	Bobsled - designed like a bobsled run -- without a fixed track. The train travels freely through a trough. *Flying* - a roller coaster ridden while parallel with the track. *Inverted* - a roller coaster which uses trains traveling beneath, rather than on top of, the track. Unlike a suspended roller coaster, an inverted roller coaster's trains are rigidly attached to the track. *Pipeline* - a coaster where riders are positioned between the rails instead of above or below. *Sit Down* - a traditional roller coaster ridden while sitting down. *Stand Up* - a coaster ridden while standing up instead of sitting down. *Suspended* - a roller coaster using trains which travel beneath the track and pivot on a swinging arm from side to side, exaggerating the track's banks and turns. *Wing* - a coaster where pairs of riders sit on either side of a roller coaster track in which nothing is above or below the riders.
Drop (ft)	Length of largest gap between high and low points of roller coaster
Length (ft)	Length of roller coaster ride
Duration (s)	Time length of roller coaster ride
Inversions?	Whether or not roller coaster flips passengers at any point (Yes or No)
# of Inversions	Number of times roller coaster flips passengers
Make	Name of the manufacturer of the roller coaster
Latitude	Latitude Coordinates, to be used for locating a roller coaster on the map
Longitude	Longitude Coordinates, to be used for locating a roller coaster on the map
Boundary	Geospatial state boundary data needed for map