

# Unit 9 - Data

## Unit Overview

Students explore and visualize datasets from a wide variety of topics as they hunt for patterns and try to learn more about the world around them from the data. Once again, students work with datasets in App Lab, but are now asked to make use of a data visualizer tool that assists students in finding data patterns. They learn how different types of visualizations can be used to better understand the patterns contained in datasets and how to use visualizations when investigating hypotheses. At the conclusion of the unit, students learn about the impacts of data analysis on the world around them and complete a final project in which they must uncover and present a data investigation they've completed independently.

## Unit Philosophy and Pedagogy

- **The Data Analysis Process:** This unit is built around a data analysis process that helps students break down how data is turned into new information about the world. Some lessons are designed around different steps of this process, like cleaning data or building visualizations. Other lessons focus on ways this process is applied in the real world in contexts like citizen science or machine learning. The data analysis process helps provide a consistent reference point as students explore the importance of data analysis in computing.
- **Exploring Data with the Data Visualizer:** The Data Visualizer is a tool built into App Lab that allows students to quickly create visualizations of the data they've added to their projects. The set of possible visualizations is intentionally limited to a few ways to change or modify the chart. The goal of this tool is to encourage the exploration of the different kinds of questions that can be answered with data visualizations, with a greater emphasis on students' ability to quickly create a variety of visualizations.

## Major Assessment and Projects

Students use the data visualizer to find and present a data story. Using what they've learned about the data analysis process, students either choose a dataset inside the data library, or upload one, of their own and create visualizations that find interesting patterns that possibly reveal new insights and knowledge. Students complete an activity guide describing their findings and the process they used in identifying them. Students will also complete an end-of-unit assessment aligned with CS Principles framework objectives covered in this unit.

## AP Connections

This unit and unit project helps build towards the enduring understandings listed below. For a detailed mapping of units to Learning Objectives and EKs please see the "Standards" page for this unit.

- DAT-2: Programs can be used to process data, which allows users to discover information and create new knowledge.
- IOC-1: While computing innovations are typically designed to achieve a specific purpose, they may have unintended consequences.

This unit includes content from the following topics from the AP CS Principles Framework. For more detailed information on topic coverage in the course review [Code.org CSP Topic Coverage](#).

- 2.3 Extracting Information from Data
- 2.4 Using Programs with Data
- 5.3 Computing Bias
- 5.4 Crowdsourcing

## Week 1

## Lesson 1: Learning from Data

App Lab

Learn how...

## Lesson 2: Exploring One Column

App Lab

Learn how...

## Lesson 3: Filtering and Cleaning Data

App Lab

Learn how...

## Lesson 4: Exploring Two Columns

App Lab

Learn how...

## Lesson 5: Big, Open, and Crowdsourced Data

Learn how big data, open data, and crowdsourcing apply the data analysis process in real world context to solve problems that matter.

## Week 2

## Lesson 6: Machine Learning and Bias

App Lab

In this lesson, students are introduced to the concepts of Artificial Intelligence and Machine Learning using the AI for Oceans widget. First students classify objects as either "fish" or "not fish" to attempt to remove trash from the ocean. Then, students will need to expand their training data set to include other sea creatures that belong in the water. In the second part of the activity, students will choose their own labels to apply to images of randomly generated fish. This training data is used for a machine learning model that should then be able to label new images on its own.

## Lesson 7: Project - Tell a Data Story Part 1

Project | App Lab

Learn how...

## Lesson 8: Project - Tell a Data Story Part 2

Project | App Lab

Learn how...

## Lesson 9: Assessment Day

Project

Students complete a multiple choice assessment which covers the unit topics.



This curriculum is available under a  
Creative Commons License (CC BY-NC-SA 4.0).

If you are interested in licensing Code.org materials for commercial purposes, **contact us**.

# Lesson 1: Learning from Data

## Overview

In this lesson students explore the Google Trends tool in order to tell a "data story" which explains both what the data shows and why that might be. Following this, students are introduced to the concept of metadata and look for the metadata of datasets on App Lab.

## Purpose

The three main purposes of this lesson are:

1. Navigating and using a real data tool (Google Trends, see below) that is external to the course
2. Getting acquainted with talking and writing about data. In particular we want to:
  - Draw a distinction between describing what the data shows and describing why it might be that way
  - In other words: describe connections and trends in data separate from drawing conclusions.
  - We want students to get in the habit of separating the what from the why when it comes to talking and writing about data
3. Introduce the concept of metadata. We want students to understand the value of data about data, that can be used to find, organize, and manage information. It also increases effective use of data by providing extra information.

## Agenda

**Warm Up (5 mins)**

**Activity (30 mins)**

**Wrap Up (10 mins)**

**Assessment: Check For Understanding**

**View on Code Studio**

## Objectives

Students will be able to:

- Differentiate between what data shows and why that might be the case
- Use Google Trends to tell a data story
- Explain the usefulness of metadata

## Preparation

- Preview Google Trends and prepare for the demo

## Links

**Heads Up!** Please make a copy of any documents you plan to share with students.

**For the Teachers**

- **CSP Unit 9 - Data** - Presentation

**For the Students**

- **Google Trends** - Link


# Teaching Guide

## Warm Up (5 mins)

### **Remarks**

Welcome to Unit 9: Data! In this unit, we are going to learn how to organize and visualize data to answer questions. We'll make charts, look for patterns, and consider the impact that data collection has on our world.

Let's get things started by asking a question.

 **Prompt:** What time of year do people tend to search online for chocolate? How could you check your guess?

### **Remarks**

Search tools are useful for efficiently finding information. How can we analyze our searches to discover patterns or trends? Let's find out!

### Discussion Goal


**Goal:** Student answers will vary, but they may include major holidays like Valentine's Day. Students may also have several different ideas for how those answers could be checked, and ultimately you will want to guide your students towards the idea of using a search tool.


## Activity (30 mins)

### **Remarks**

When you post information to a social network, watch a video online, or simply search for information on a search engine, some of that data is collected, and you reveal what topics are currently on your mind. When a topic is quickly growing in popularity, it is often said to be trending, but there are many different trends or patterns we might find in this data, including historical trends. These patterns might help us to identify, understand, and predict how our world is changing.

We will be using Google Trends which is a tool that allows you to visualize data about search history across different times and locations. We can use it to look for interesting patterns, trends, or relationships between multiple trends and try to tell the story that pattern is showing.

 **Display:** The chart relates to the warm-up question and shows searches for chocolate over the past 5 years.

 **Prompt:** What pattern do you notice? What could be the reason for that pattern? Does this data support your earlier guesses?


### **Remarks**


This chart gives us information: a collection of facts and patterns extracted from data. We can use this information to identify trends, make connections and address problems.

 There are two distinctions we need to make when looking at a chart or visualization:

- What does the data show?
- Why might that be the case?

The "what" is the facts of the matter. Chocolate searches spike in December. The "why" is an informed opinion. Chocolate searches spike around the time of the winter holidays because it's a common gift.

 **Do This:** Demonstrate how to use Google Trends in front of the class. Add several search items (for example: streaming, mp3, cd) or take suggestions from the class. Demonstrate usage of the dropdowns to narrow focus. As a class look for patterns in the data. Practice stating the *What* and the *Why* as you make charts with the class.

 **Do This:** Now students use Google Trends themselves to look for "data stories". Depending on class time, allow students to share their data stories with partners or with the class as a whole.

### **Remarks**

Using Google Trends to find patterns represents an important concept in understanding how programs are used to gain insight and knowledge from data. When you worked with the tool, there were elements of both interaction and iteration. You interacted with the tool trying out several different options, and repeated the process (iteration)

until you came to a possible conclusion.

As you were determining the **Why** in your data stories, you may have been tempted to draw concrete connections. However, it's important to remember that correlation (similarities, patterns) does not equal causation (this thing caused that thing). There can be any number of reasons why a pattern or interesting data point may appear in a chart - and our job is to make an informed decision while recognizing that there may be multiple factors at play. Usually additional research with several data sets is necessary to understand the exact nature of the relationship between data. Did this one thing cause another thing to happen? Do more people search for chocolate because they want to give it as a gift, or could there perhaps be another reason?

In this unit we will be making charts to help answer questions:

- "I think this visualization tells me this..."
  - Something is more popular than something else
  - Something is more important than something else
  - Something has become more or less searched over time
- "... but I'm not sure because..."
  - I don't know exactly how the data was collected
  - This might tell me people searched for green more than red, but it doesn't tell me why they do that or that green is a better color
  - We need more data!

When we consider datasets, it's helpful to know as much about those datasets as possible. Where did the data come from? How much data is included? When was it collected?

All of this information is considered **metadata** which is defined as "data about data".

We can have metadata about any digital data. For example, this picture contains metadata that tells us the when the picture was created, what the resolution is, and how many people have downloaded it. Datasets can also have metadata that explains more about the information in the dataset. You've seen this in App Lab!

**Prompt:** What is the metadata for the chart you created in Google Trends?

### Remarks

Now let's explore metadata in App Lab in the datasets you've used in previous units. Where can we find this information? What can it tell us about the data? How can it help us organize the data?

**Do This:** Students navigate to Level 2 on Code

Studio where they open the data tab and look at the metadata for a table. Then they should share with a partner where the data comes from and what they can learn from the metadata.

### Discussion Goal

**Goal:** Students may bring up the sources of the data (Google Searches) or the dropdowns that control how the chart is built. If students want to dig deeper into the metadata, consider pointing them to the **Google Trends Help Page**.

## Wrap Up (10 mins)

**Review:** Review key takeaways on the slide. Students may want to jot down notes in their journals.

**Journal:** Students add the following word and definition to their journal: metadata.

## Assessment: Check For Understanding

Check For Understanding Question(s) and solutions can be found in each lesson on Code Studio. These questions can be used for an exit ticket.

**Question:** Below is an image from Google Trends that plots Cats and Dogs. Choose the most accurate description of what this data is actually showing based on what you know about how Google Trends works.

## Standards Alignment

CSTA K-12 Computer Science Standards (2017)

- ▶ **DA** - Data & Analysis

CSP2021

- ▶ **DAT-2** - Programs can be used to process data



This curriculum is available under a  
Creative Commons License (CC BY-NC-SA 4.0).

If you are interested in licensing Code.org materials for commercial purposes, **contact us**.

# Lesson 2: Exploring One Column

## Overview

Students will practice making conclusions from charts and learn to use the Data Visualizer in App Lab to make create two different kinds of charts a bar chart, and a histogram. The lesson begins with a quick prompt to review the reasons charts are useful for looking at data. Students then practice reading a bar chart and review the kinds of questions it is and is not useful for answering. Afterwards they build different bar charts in the Data Visualizer and discuss why some are or are not useful. Afterwards they learn how to make histograms for building charts in instances where bar charts may not be useful. Students record their work on an activity guide. The lesson concludes with a brief review of what they learned.

## Purpose

This lesson introduces students to the Data Visualizer tool in App Lab as well as two important chart types, a bar chart, and a histogram. Students will build skills in using this tool and reading data that they will continue to develop throughout the unit.

In the Wrap Up of this lesson students are introduced to the Data Analysis Process, a set of steps that can be used to use data to answer questions. Students will continue to refer to this process throughout the remainder of the unit.

## Agenda

**Warm Up (5 mins)**

**Activity (35 mins)**

Reading a Bar Chart

Making Bar Charts

Reading a Histogram

**Wrap Up (5 mins)**

**Assessment: Check For Understanding**

[View on Code Studio](#)

## Objectives

Students will be able to:

- Create a bar chart and a histogram in App Lab's data visualizer
- Draw conclusions by reading bar charts and histograms
- Explain the reasons that someone would create either a bar chart or a histogram in order to explore a single column of data

## Preparation

- Try using the Data Visualizer yourself to make some of the charts students will create as part of completing the lesson

## Links

**Heads Up!** Please make a copy of any documents you plan to share with students.

**For the Teachers**

- [CSP Unit 9 - Data](#) - Presentation

**For the Students**


- [Exploring One Column](#) - Activity Guide

[Make a Copy](#) ▾




# Teaching Guide

## Warm Up (5 mins)


 **Prompt:** Why do people make visualizations out of data?

**Discuss:** Have students brainstorm silently on their own, then have them share with neighbors, and finally have them share out with the room.

### **Remarks**

 When looking at the datasets on App Lab, the table format may be familiar to you. You may have seen data laid out and organized in a spreadsheet. Once the data is organized, we can now find trends and patterns. This is where visualizations are useful!

People make visualizations for a number of reasons. They help humans see lots of data at once and help us find patterns that might otherwise be "invisible". Today we're going to learn how to make two new types of visualizations.

 **Display:** Display the slide showing the Data Analysis Process. Students will continue to refer to this throughout the unit, including at the end of today's lesson. Simply make students aware of it as a framework and then quickly move on to the main activity.

### Discussion Goal

**Goal:** This discussion should be a quick prompt to kick off today's activity. Students spent the previous class making visualizations and looking at data sets, but they didn't think about how to make visualizations with the data. Try to quickly guide conversation towards some of the takeaways on the following slide, then move on to the day's activity.

## Activity (35 mins)

### Reading a Bar Chart


 **Prompt:** Which of these questions does this chart answer? Be ready to discuss with a classmate.


**Discuss:** Run a short discussion, asking students to share which of the five questions they believe could be answered with the chart. Students should hopefully notice:


- Only question 1 and 3 could be answered by the chart
- They may have additional questions about how the chart was created or what it means. Hold those questions for the following slide

 **Display:** Show the following slide introducing bar charts and how they are used

### Making Bar Charts

 **Distribute:** Give students access to **Exploring One Column - Activity Guide**, ideally in digital form.

 **Do This:** Have students go to the data visualizer and recreate the chart they just examined together. Use this as an opportunity to make sure every student understands how to use the visualizer.


 **Do This:** Have students complete side one of the activity guide, making a bar chart for each column in the Dogs table and deciding if it is useful or not. They will also need to paste in one chart they find is useful and answer a pair of prompts.

 **Discuss:** Optionally discuss the different columns that students chose as useful.

### Teaching Tip

**Fill Out the Activity Guide Digitally:** The Data Visualizer allows students to quickly copy and paste visuals into documents and slide shows. If at all possible in your classroom environment, have students complete this activity guide digitally.

### Reading a Histogram

 **Display:** Show the slide explaining why some bar charts are not useful.

**Prompt:** In the Visualizer make a "Histogram" for "Max Weight" with a "Bucket Size" of 20.

- Which of these makes it easier to understand the data?
- What do you think the "histogram" is doing to visualize the data differently?

**Discuss:** Discuss how the histogram seems to be displaying the data differently from the bar chart.

**Display:** Show the slide explaining what a histogram is useful for

**Prompt:** Have students complete side 2 of their activity guide.

**Discuss:** Optionally have students share out the results of their work on side 2 of the activity guide.

## Wrap Up (5 mins)

**Prompt:** Which steps of the Data Analysis Process did you see in today's activity? Where did you see them?

**Discuss:** Have students share their reflections and lead a quick discussion about where they saw each step.

**Review:** Review the key takeaways from the lesson

## Assessment: Check For Understanding

*Check For Understanding Question(s) and solutions can be found in each lesson on Code Studio. These questions can be used for an exit ticket.*

**Question:** Why would someone make a histogram instead of a bar chart?

### Discussion Goal

**Goal:** Use this discussion for two purposes. First, use it to build a shared comfort with what a bar chart shows. Have students share out which charts they think are useful and what pieces of information they read from them. Remind students that bar charts are particularly good at finding how common a value is in a column. The second goal is to prepare students to see the value of a histogram. Many bar charts are not at all useful because they have too many unique values. Using a histogram to group these values can help make more useful charts.

### Discussion Goal

**Goal:** Use this discussion to start exploring the differences between a bar chart and a histogram. For example, students should note that there seem to be "fewer but wider" bars in the histogram. They should also notice that values are sorted correctly. This makes the chart overall easier to read.

### Discussion Goal

**Goal:** Use this discussion to have students share out how they are designing histograms based on the columns in the dogs data set. Things to look for:

- Students are only creating histograms for numeric columns
- Students are picking columns that have a number of unique values
- Students are picking good bucket sizes to group their data in a way that makes it possible to interpret
- Students are correctly reading charts to understand what they are showing

## Discussion Goal

**Goal:** This is the first opportunity for students to talk through the Data Analysis Process, a framework they will return to many times throughout this course. Use this discussion to review some key points

- Students in today's lesson may not have "cleaned and processed" data but they should've seen every other step.
- Students answered questions and carefully made conclusions
- Students created visualization and used them to find patterns.
- Students chose the datasets that would best answer their question. In other instances they might've gone off to find the data themselves but the Data Library in App Lab helps avoid them needing to do that.

## Standards Alignment

CSTA K-12 Computer Science Standards (2017)

- ▶ **DA** - Data & Analysis

CSP2021

- ▶ **DAT-2** - Programs can be used to process data



This curriculum is available under a Creative Commons License (CC BY-NC-SA 4.0).

If you are interested in licensing Code.org materials for commercial purposes, **contact us**.

# Lesson 3: Filtering and Cleaning Data

## Overview

In this lesson, students explore the challenges of working with a messy dataset. First students learn how to identify issues using the Data Visualizer, and then manually clean the data. Following this, students learn about the filtering tools in the Data Visualizer, and use a guided activity to answer data questions that require filtering a dataset.

## Purpose

The goal of this lesson is to introduce two crucial concepts when working with data: cleaning and filtering. The datasets in App Lab are generally clean, so a "messy" dataset is imported to a level for students to explore. After discovering why datasets need to be cleaned, students manually clean the dataset.

When working with a dataset to answer a question, the user may want to focus on a subset of the data. In this lesson, students are introduced to the filtering tools in the Data Visualizer so they can accurately filter for the information they are looking for.

## Agenda

**Warm Up (2 mins)**

**Activity (33 mins)**

**Wrap Up (10 mins)**

**Assessment: Check For Understanding**

[View on Code Studio](#)

## Objectives

Students will be able to:

- Explain why data needs to be cleaned
- Use the Data Visualizer to filter data
- Create filtered charts that answer specific questions

## Preparation

- Preview the filter tool in the Data Visualizer
- Prepare for the demo in the Activity

## Links

**Heads Up!** Please make a copy of any documents you plan to share with students.

For the Teachers

- [CSP Unit 9 - Data](#) - Presentation

For the Students

- [Filtering Data Unit 9 Lesson 3](#) - Activity Guide [Make a Copy](#) ▾

# Teaching Guide

## Warm Up (2 mins)

### **Remarks**

We've started to explore how to use charts to process data stored in a table, but there are challenges with doing this. The ability to process that data depends on the users and available tools. Today, we are going to explore ways to refine the data the we can answer even more questions using the Data Visualizer.

 **Display:** Set the scene for the first activity.

### **Remarks**



Imagine you have used a survey to collect information from students. This is aligned with the first step in the Data Analysis Process.

All of that data is now stored in a table. You are excited to dig into the data and see what you can learn. Let's go!

## Activity (33 mins)

**Do This:** Instruct students to navigate to Level 2 on Code Studio.

- Open the data tab
- Familiarize yourself with the imported table
- Open the Data Visualizer
- Make Charts:
  - Average Hours of Sleep
  - Favorite Subject

  **Discuss:** With partners, students discuss the following prompts:


- What problems came up when trying to create these charts?
- What problems do you see in the data?

### Discussion Goal

**Goal:** Students should point out some issues like 4 and "four" being charted as different values.

### **Remarks**

Datasets can bring about challenges, no matter what their size. There can be incomplete data and invalid data. You might want to combine two tables, with inconsistent data. All of this requires data to be cleaned.

 When does data need to be cleaned?

- Data is incomplete
- Data is invalid
- Multiple tables combined into one

What leads to "messy data"?

- Users enter in different types of data ("two", 2)
- Users use different abbreviations to represent the same information ("February", "Feb", "Febr")
- Data may have different spellings ("color", "colour") or inconsistent capitalization ("spring", "Spring")

 When we clean data, the goal is to make it uniform without changing meaning.

For example: "two" can be changed to 2

 **Do This:** With a partner:

- Clean the Student Info table
- Look for:
  - Different types of data ("two", 2)
  - Different abbreviations to represent the same information ("February", "Feb", "Febr")

- Different spellings ("color", "colour")
- Inconsistent capitalization ("spring", "Spring")
- Manually update cells with messy data so they are consistent with other cells, while not changing the meaning of the data.

**Do This:** Once students have finished cleaning their data, they should remake the original charts and compare with others.

### **Remarks**

Your charts should look similar to others, but it depends how you cleaned your data.

Now you are able to accurately visualize the data in your table!

For this activity, we cleaned a dataset that is similar to one you might create yourself or have uploaded from another location. However, the datasets in the dataset library have already been cleaned for you!

**Prompt:** What if I only wanted to look at a subset of my data? How could I do this? For example: I only want to investigate dogs with low shedding.

### **Remarks**

The best way to look at a subset of data is to use a filter. In Unit 5, we filtered data programmatically using traversals in order to gain insight into knowledge from that data.

Software programs with built-in tools (like the Data Visualizer) can also be used to filter data. These tools help us find specific information in the data and look for patterns.

### **Discussion Goal**

**Goal:** We are building towards the idea of filtering data. Students filtered programmatically in previous units, but don't yet know how to use the Data Visualizer tool to accomplish the same goal.

In the discussion, students may suggest deleting unwanted rows or creating new tables. Challenge students by asking what they would do if they wanted to look at several different subsets in the same data? Is it reasonable to create a new table for each subset?

**Do This:** Demonstrate how to filter data for the class.

1. Open up Level 3.
2. Discuss that you want to find out about the minimum life span for low shedding cats. What will you filter by? Minimum life span or shedding?
3. We will filter by shedding. This means the data for cats who have low shedding is the only data that will be shown.
4. Select "Bar Chart" and "Min Life Span" from the dropdowns.
5. Discuss what the chart displays: Cats who are low shedding have a minimum life span of 10-14 years, with 12 being the most common.

### **Remarks**

When filtering, the most challenging part is deciding what value you will filter by. Think to yourself: what's the limiting factor? What do I want to make sure all cats, dogs, etc. have to be included in this subset?

It's time for you to play around in the tool yourself!

**Distribute:** Share the Activity Guide with students. Alternatively, they can click on the link in Level 1. This Activity Guide is designed to be used digitally - do not print.

**Do This:** For the rest of the activity time, students work through the Activity Guide filtering two datasets and answering questions. Students need to copy/paste their charts into the Activity Guide.

## **Wrap Up (10 mins)**

**Prompt:** Why is "Clean and/or Filter" an important part of the Data Analysis Process? What are situations when you would filter vs. clean your data?

### **Remarks**

Great job filtering and cleaning data today! This is a useful skill that will help you create more powerful visualizations.

There are other things that programs can do to better prepare data for visualizations including combining datasets, clustering data, and classifying data. The Data Visualizer isn't the best tool for these more advanced concepts, so if you are interested in exploring more data analysis we suggest researching tools like spreadsheets.

#### Discussion Goal

**Goal:** Messy data can lead to confusing charts that could be misinterpreted. Data should be filtered when you want to focus on a subset of the data. Filtering allows you to return to the full dataset if you have unanswered questions.

## Assessment: Check For Understanding

*Check For Understanding Question(s) and solutions can be found in each lesson on Code Studio. These questions can be used for an exit ticket.*

**Question:** What makes manually cleaning data challenging?

## Standards Alignment

CSTA K-12 Computer Science Standards (2017)

► **DA** - Data & Analysis

CSP2021

► **DAT-2** - Programs can be used to process data



This curriculum is available under a Creative Commons License (CC BY-NC-SA 4.0).

If you are interested in licensing Code.org materials for commercial purposes, **contact us**.

# Lesson 4: Exploring Two Columns

## Overview

Students will practice making conclusions from charts and learn to use the Data Visualizer in App Lab to make create two different kinds of charts, a cross tab, and a scatterplot. Students will practice reading each type of chart before learning to make them in the Data Visualizer. Students will track their work using a provided activity guide. The lesson concludes with a review of key takeaways.

## Purpose

This lesson introduces students to two new ways of make visualizations in the Data Visualizer. The crosstab and scatter chart are new in that they allow students to see patterns across multiple variables, noticing how one might seem to change (or correlate) with another. This is good preparation for their unit project in which they'll need to make and interpret a data visualization of their own.

## Agenda

**Warm Up (5 mins)**

**Activity (30 mins)**

Reading Crosstab Charts

Making Crosstab Charts

Reading Scatter Charts

Making Scatter Charts

**Wrap Up (10 mins)**

**Assessment: Check For Understanding**

**View on Code Studio**

## Objectives

Students will be able to:

- Create a crosstab and scatter charts in App Lab's data visualizer
- Draw conclusions by reading crosstab and scatter charts
- Explain the reasons that someone would create either a crosstab and scatter chart in order to explore two columns of data

## Preparation

- Review how to make crosstab and scatter charts in the Data Visualizer

## Links

**Heads Up!** Please make a copy of any documents you plan to share with students.

For the Teachers

- **CSP Unit 9 - Data** - Presentation

For the Students


- **Exploring Two Columns** - Activity Guide

Make a Copy ▾




# Teaching Guide

## Warm Up (5 mins)

 **Prompt:** Imagine you wanted to know which hour of the day you and your classmates are happiest. What kind of data would you collect? How do you think you'd analyze it?

**Discuss:** Have students brainstorm silently on their own, then have them share with neighbors, and finally have them share out with the room.

### **Remarks**


 We just heard a lot of really interesting ideas here and they should make for some fascinating analysis. The main thing I noticed though, is that we need at least two different pieces of information. The time of day, and how happy people are. That means we're going to need to start thinking about ways to analyze more than one column of data. Today we're going to start looking at two different ways to do that

### Discussion Goal

**Goal:** This discussion should preview the idea that students will need to collect multiple pieces of information about both the time of day and how happy people are. Beyond highlighting that fact there's no particular direction this conversation needs to go.


## Activity (30 mins)

### Reading Crosstab Charts

 **Prompt:** How many "Herding" breeds live a maximum of 12 years? What is the most common maximum life span for "Working" breeds? Which breed group lives the shortest? Which breed group lives the longest? How do you know? How confident are you in your answers?

**Discuss:** Run a short discussion and help students read the chart together. They will hopefully note the following:


- 4 herding breeds live a maximum of 12 years
- The most common maximum lifespan of working breeds is 12 years
- Toy breeds seem to live the longest
- Working breeds seem to be the shortest
- Students can see the spread of life expectancies for different breeds but it's not cut and dry.


 **Display:** Show the following slide introducing cross tab charts and how they work.

### Teaching Tip

**Understanding Crosstab:** Give students some time to think and discuss why a crosstab chart might be a good choice for finding patterns like the ones indicated on this slide. Further reinforce the fact that if either column has too many values you may end up with an enormous chart.

### Making Crosstab Charts

 **Distribute:** Give students access to **Exploring Two Columns - Activity Guide**, ideally in digital form.

 **Prompt:** Have students go to the data visualizer and recreate and interpret the two charts requested for both "Words" and the "Favorite Classes" data set.

**Discuss:** Optionally have students share their progress and discuss the conclusions they reached on side one of the activity guide.

### Teaching Tip

**Fill Out the Activity Guide Digitally:** The Data Visualizer allows students to quickly copy and paste visuals into documents and slide shows. If at all possible in your classroom environment, have students complete this activity guide digitally.

### Reading Scatter Charts

 **Prompt:** Is there a pattern? How can you tell?

**Discuss:** Have students share whether they think there's a pattern between the order in which states were admitted and their size.

**Display:** Show the following slide introducing scatter charts and how they work.

#### Discussion Goal

**Goal:** Students will hopefully notice later states are relatively larger than earlier added ones. The trend does not necessarily reflect any causation, but there is a slight uptick in state size as more are added.

## Making Scatter Charts

**Prompt:** Have students fill in side 2 of their activity guides practicing making scatter charts.

**Discuss:** Have students share their responses and talk through their conclusions.

## Wrap Up (10 mins)

**Review:** Review the three slides of the key takeaways and hit on the following points.

- Students know a lot of different ways to find patterns in data
- Review the chart showing how they can decide the type of visualization to create
- Have students record key takeaways in their journal

## Assessment: Check For Understanding

*Check For Understanding Question(s) and solutions can be found in each lesson on Code Studio. These questions can be used for an exit ticket.*

**Question:** How are the questions you can investigate with scatter or crosstab charts different from the ones you can investigate with bar charts or histograms?

## Standards Alignment

CSTA K-12 Computer Science Standards (2017)

- ▶ **DA** - Data & Analysis

CSP2021

- ▶ **DAT-2** - Programs can be used to process data



This curriculum is available under a Creative Commons License (CC BY-NC-SA 4.0).

If you are interested in licensing Code.org materials for commercial purposes, **contact us**.

# Lesson 5: Big, Open, and Crowdsourced Data

## Overview

Students will complete a jigsaw of three different topics at the intersection of data, computing, and global impacts. These are topics, big data, crowdsourcing, and open data. Students will watch videos or listen to audio recordings about the different topics. Groups will each complete an activity guide about their topic before having individuals from each group share out their findings. The lesson concludes with a review of key points.

## Purpose

This lesson zooms back out from the data analysis process to the ways that is applied in a wide variety of contexts. Students learn how big data, open data, and crowdsourcing apply this process in interesting ways that cleverly modify this process. For a summary of key points of this lesson review the key takeaways in the slides. In short however:

- Big data: "Collect huge amounts of data so we can learn even more from it"
- Open data: "sharing data with others so they can analyze it"
- Crowdsourcing: "collecting data from others so you can analyze it"

This lesson further builds towards the following lesson on machine learning which explores a different application of the data analysis process.

## Agenda

**Warm Up (5 mins)**

**Activity (30 mins)**

**Wrap Up (5 mins)**

**Assessment**

[View on Code Studio](#)

## Objectives

Students will be able to:

- Define and explain the impacts of crowdsourcing, crowdfunding, and citizen science
- Explain why in some contexts large amounts of data need to be analyzed in parallel and scalable systems
- Explain the impact of open data on scientific research and discovery

## Preparation

- ☐ Ensure students will be able to access all of the videos / articles linked in the lesson.
- ☐ Review at least the key takeaways and ideally some of the content from each topic to ensure you understand how these topics relate to what students have studied in previous lessons.

## Links

**Heads Up!** Please make a copy of any documents you plan to share with students.

For the Teachers


- [CSP Unit 9 - Data](#) - Presentation

For the Students

- [Big, Open, and Crowdsourced Data](#) - Activity Guide [Make a Copy](#)

# Teaching Guide

## Warm Up (5 mins)

 **Prompt:** With a partner review the data analysis process and for each step talk through:

- What is this step and why is it important?
- Where have we done this step together?
- What could go wrong if you do this step poorly.


**Discuss:** Have students brainstorm silently on their own, then have them share with neighbors, and finally have them share out with the room.


### **Remarks**


Today we're going to be looking at a lot of ways that data is being used in exciting and innovative ways. We're going to stop looking just at the data in App Lab and start thinking about the impacts data has on our lives. Along the way we'll talk about how the data analysis process looks different or has been manipulated in different contexts in order to answer questions or make decisions that matter.

## Activity (30 mins)

**Group:** Place students in pairs

 **Distribute:** Give each pair a copy of the **Big, Open, and Crowdsourced Data - Activity Guide**

 **Prompt:** With a partner *Choose one of the topics* Watch the related videos / listen to the podcasts \* Take notes and be ready to share responses to the questions on your activity guide

 **Discuss:** Have members from each topic share the conclusions from their watching and research. Make sure that students from each group have time to share

- What the topic is
- The key vocabulary they were responsible for researching
- How this concept uses or modifies the data analysis process
- Examples of the problems this technique is being used to solve

## Wrap Up (5 mins)

 **Review:** Review key takeaways on the slides

## Assessment

**Assess:** You can collect and evaluate students' activity guides

### Discussion Goal

**Goal:** This is designed entirely to be a review of concepts students have previously covered and doesn't foreshadow the lesson of the day. If you feel confident your students are comfortable with this process already then quickly move on to the main activity of the day.

### Teaching Tip

**Complete the Activity Digitally:** Students will have a much easier time accessing articles and videos if they complete the activity digitally. Alternately students can complete printed versions of the activity guide but still access links through the digital versions.

**Supporting the Jigsaw:** In this lesson students do a jigsaw of a number of different topics. Students will need access to computers and should spend roughly 10 minutes in each group listening to audio / video content. During this period circulate the room encouraging them to focus on the questions they've been asked to respond to. This will also help you anticipate or even specifically ask different students to participate during the discussion.

# Standards Alignment

CSTA K-12 Computer Science Standards (2017)

- ▶ **DA** - Data & Analysis

CSP2021

- ▶ **DAT-2** - Programs can be used to process data
- ▶ **IOC-1** - While computing innovations are typically designed to achieve a specific purpose, they may have unintended consequences



This curriculum is available under a  
Creative Commons License (CC BY-NC-SA 4.0).

If you are interested in licensing Code.org materials for commercial purposes, **contact us**.

# Lesson 6: Machine Learning and Bias

## Overview

In this lesson, students are introduced to the concepts of Artificial Intelligence and Machine Learning using the AI for Oceans widget. First students classify objects as either "fish" or "not fish" to attempt to remove trash from the ocean. Then, students will need to expand their training data set to include other sea creatures that belong in the water. In the second part of the activity, students will choose their own labels to apply to images of randomly generated fish. This training data is used for a machine learning model that should then be able to label new images on its own.

## Purpose

This tutorial is designed to quickly introduce students to machine learning, a type of artificial intelligence. Students will explore how training data is used to enable a machine learning model to classify new data.

## Agenda

**Warm Up (5 mins)**

**Activity (35 mins)**

**Wrap Up (5 mins)**

**Assessment: Check For Understanding**

**View on Code Studio**

## Objectives

Students will be able to:

- Train and test a machine learning model.
- Reason about how human bias plays a role in machine learning.

## Preparation

- Review and complete the online tutorial yourself. If you are not going to use AI for Oceans, explore the other options listed below.

## Links

**Heads Up!** Please make a copy of any documents you plan to share with students.

For the Teachers

- **CSP Unit 9 - Data** - Presentation

# Teaching Guide

## Warm Up (5 mins)


 **Prompt:** How can machines "learn"?

**Discuss:** Have students brainstorm silently on their own, then have them share with neighbors, and finally have them share out with the room.

### **Remarks**


Today we're going to be learning more about Machine Learning and its impacts.


## Activity (35 mins)


 **Video:** Play the video "What is Machine Learning".


### **Remarks**

Machine learning refers to a computer that can recognize patterns and make decisions without being explicitly programmed. In this activity you're going to supply the data to train your own machine learning model. Imagine an ocean that contains creatures like fish, but also contains trash dumped by humans. What if we could train a computer to tell the difference and then use that technology to help clean the ocean?


 **Do This:** Direct students to Levels 3-5 on Code Studio. Students should spend around five minutes total on these levels. Prompt their thinking with the "Consider" on the slide. To program A.I., use the buttons to label an image as either "fish" or "not fish". Each image and label becomes part of the data used to train A.I. to do it on its own. Once trained, A.I. will attempt to label 100 new images on its own, then present a selection that it determined have the highest probability of being "fish" based on its training. Students who consistently label things correctly should see an ocean full of different types of sea creatures, without much (or any) other objects.

 **Discuss:** How well did A.I. do? How do you think it decided what to include in the ocean?

 **Video:** Play the video "Training Data & Bias".

 **Prompt:** How do you think your training data influenced the results that A.I. produced?

### **Remarks**

 In the second half of the activity, you will teach A.I. about a word of your choosing by showing it examples of that type of fish. As before, A.I. doesn't start with any training data about these labels. Even though the words in this level are fairly objective, it's possible that you will end up with different results based on their training data. You might even intentionally train A.I. incorrectly to see what happens!

### Discussion Goal

**Goal:** Based on yesterday's conversations, answers may vary. Steer the discussion towards conversations around the role that humans play in machines learning. It's ok if the discussion here is short - you are setting the stage for the upcoming activity.

### Teaching Tip

**Alternatives to AI For Oceans:** AI for Oceans was originally developed as an Hour of Code activity that can be completed by students with any device available. We have modified it for its usage here. Depending on your classroom situation, you might opt to replace the activity with:

- **Teachable Machines** - *Teachable Machine is a web-based tool that makes creating machine learning models fast, easy, and accessible to everyone. Teachable Machine is flexible - use files or capture examples live. It's respectful of the way you work. You can even choose to use it entirely on-device, without any webcam or microphone data leaving your computer.*
  - If your classrooms devices have cameras, Teachable Machines offers an engaging way to create training sets. Encourage students to teach the machine to represent rock, paper, or scissors with hand gestures. What are some possible ways for bias to enter in?
- **Machine Learning for Kids** - *This free tool introduces machine learning by providing hands-on experiences for training machine learning systems and building things with them. It provides an easy-to-use guided environment for training machine learning models to recognise text, numbers, images, or sounds.*
  - Machine Learning for Kids is a great option if your students want to work with text samples. Teach the machine to recognize words or passages that are happy or sad. Lots to play around with here!

**Do This:** Direct students to Levels 7-8 on Code Studio. Students should spend around five minutes total on these levels. Prompt their thinking with the "Consider" on the slide. Here, as before, students will use training data to teach A.I. to recognize different types of fish. The words in this list are intentionally more subjective than what students will have seen so far. Encourage students to decide for themselves what makes a fish look "angry" or "fun". Two students may choose the same label and get a very different set of results based on which fish traits were their focus. Encourage students to discuss their findings with each other or go back and choose new words. Each student will rely on their own opinions to train A.I. which means that A.I. will learn with the same biases held by the students. As students begin to see the role their opinion is playing, ask them to reflect on whether this is good or bad, and how it might be addressed.

**Prompt:** How could biased data result in problems for artificial intelligence? What are ways to address this?

**Video:** Play the video "How I'm fighting bias in algorithms" with Joy Buolamwini.

**Prompts:**

- How can computing innovations which make use of Machine Learning reflect existing human bias?
- How could it be used to discriminate against groups of individuals?
- How can that bias be minimized?

**Remarks**

As we've seen, problems of bias are often created by the type or source of data being collected. Collecting more data does not mean that the bias is removed. Computing innovations can reflect existing human biases because of biases written into the algorithms or biases in the data used by the innovation.

Machine learning and data mining have led to innovations in medicine, business, and science but information discovered in this way has been used to discriminate against groups of individuals.

Programmers (that includes you!) should take action to reduce bias in algorithms used for computing innovations as a way to combat existing human biases. Be on the lookout! Bias can occur at any level in software development.

**Review:** Play the video "Impact on Society" which recaps the concepts discussed today.

## Wrap Up (5 mins)

**Prompt:** Which steps of this process do you think have to be done by humans? Would you be concerned if any of them were automated?

**Discuss:** Time may be running short at this point in the class. Encourage students to share with a neighbor or share out with the room. The conversation should focus around bias.

### Content Corner

Every image in this part of the tutorial is fed into a neural network that has been pre-trained on a huge set of data called **ImageNet**. The database contains over 14 million hand-annotated images. ImageNet contains more than 20,000 categories with a typical category, such as "balloon" or "strawberry", consisting of several hundred images. When A.I. is scanning new images and making its own predictions in the tutorial, it is actually comparing the possible categories for the new image with the patterns it found in the training dataset.

### Discussion Goal

**Goal:** Get students to reflect on their experience so far. It is important at this point that they realize the labeling they are doing is actually programming the computer. The examples they show A.I. are the "training data".

### Content Corner

The fish in this tutorial are randomly generated based on some pre-defined components, including mouths, tails, eyes, scales, and fins, with a randomly chosen body color, shape, and size. Rather than looking at the actual image data, A.I. is now looking for patterns in these components based on how the student classifies each fish. It will be more likely to label a fish the same way the student would have if it has matching traits.

### Discussion Goal

**Goal:** At this point, students should have some preliminary thoughts on how biased data leads to problems for artificial intelligence. They may bring up that if the data sets are trained incorrectly, there will be incorrect or misinterpreted conclusions. It can be addressed through diverse training sets. The following video dives into this subject further.



## **Remarks**

At this point, you've fully explored the core parts of the Data Analysis Process. Ultimately you are able to use the new information gained through visualizing and finding patterns (whether yourself or using Machine Learning) to make decisions. This is why being careful about bias is so important!

## Assessment: Check For Understanding

*Check For Understanding Question(s) and solutions can be found in each lesson on Code Studio. These questions can be used for an exit ticket.*

**Question:** Think about examples of Machine Learning you may have encountered in the past such as a website that recommends what video you may be interested in watching next. Are the recommendations ever wrong or unfair? Give an example and explain how this could be addressed.

## Standards Alignment

CSTA K-12 Computer Science Standards (2017)

- ▶ **AP** - Algorithms & Programming

CSP2021

- ▶ **DAT-2** - Programs can be used to process data
- ▶ **IOC-1** - While computing innovations are typically designed to achieve a specific purpose, they may have unintended consequences



This curriculum is available under a Creative Commons License (CC BY-NC-SA 4.0).

## Teaching Tip

You can share these stories with your class to help them see how AI will impact the future.

- **Food Waste Is a Serious Problem. AI Is Trying to Solve It**
- **AI tech can identify genetic disorders from a person's face**
- **How an AI Startup Designed a Drug Candidate in Just 46 Days**
- **MIT AI tool can predict breast cancer up to 5 years early**
- **The Army steps up its pace on self-driving cars**
- **San Francisco says it will use AI to reduce bias when charging people with crimes**
- **AI is helping scholars restore ancient Greek texts on stone tablets**

If you are interested in licensing Code.org materials for commercial purposes, **contact us**.

# Lesson 7: Project - Tell a Data Story Part 1

## Overview

First day of a project where students use the Data Analysis Process to tell a data story.

## Purpose

The goal of this lesson is for students to put into use all of the data analysis skills they have practiced throughout this unit.

## Agenda

**Warm Up (2 mins)**

**Activity (43 mins)**

**Wrap Up (0 mins)**

**View on Code Studio**

## Objectives

Students will be able to:

- Follow the Data Analysis Process to tell a data story
- Write a short explanation of a data set referencing the metadata
- Create an effective visualization

## Preparation

- Preview the Project Guide

## Links

**Heads Up!** Please make a copy of any documents you plan to share with students.

**For the Teachers**

- **CSP Unit 9 - Data** - Presentation

**For the Students**

- **CSP U9 Project Guide - Tell a Data Story - Activity Guide**

# Teaching Guide


## Warm Up (2 mins)

### **Remarks**

For the next few days you will work on a project that tells a data story using the Data Analysis Process. Let's get started!

## Activity (43 mins)

**Distribute:** Make sure that all students have access to the Project Guide. Students can find this on the first level of Lesson 7 on Code Studio.

 **Do This:** Read through the Project Guide as a class. Students are focusing on Page 1 of the guide today, which covers the first three parts of the Data Analysis Process.

## Wrap Up (0 mins)

**Note:** There is no wrap up today. Allow the full time for students to work on their projects.

### Teaching Tip

Choosing a dataset for the project and creating the visualization may be a cyclical process. Encourage students to explore the datasets using the Data Visualizer before making a firm decision. Allow a good amount of time for this exploration.

Students should look for visualizations that lead to a compelling narrative. This will result in a more compelling and insightful written response.

Student answers on the Project Guide will vary. You might ask students to write full paragraphs or allow bullet points. You are encouraged to modify for your classroom environment.

## Standards Alignment

CSTA K-12 Computer Science Standards (2017)

► DA - Data & Analysis



This curriculum is available under a Creative Commons License (CC BY-NC-SA 4.0).

If you are interested in licensing Code.org materials for commercial purposes, **contact us**.

# Lesson 8: Project - Tell a Data Story Part 2

## Overview

Second day of a project where students use the Data Analysis Process to tell a data story.

## Purpose

The goal of this lesson is for students to put into use all of the data analysis skills they have practiced throughout this unit.

## Agenda

**Warm Up (0 mins)**

**Activity (35 mins)**

**Wrap Up (10 mins)**

**Assessment: Project**

**View on Code Studio**

## Objectives

Students will be able to:

- Follow the Data Analysis Process to tell a data story
- Explain information in a visualization
- Describes new insights or decisions that can be made based on a visualization
- Recognize and explain potential bias in a dataset or interpretation

## Preparation

- ▣ Review the rubric in the Project Guide

## Links

**Heads Up!** Please make a copy of any documents you plan to share with students.

For the Teachers

- **CSP Unit 9 - Data** - Presentation

For the Students

- **CSP U9 Project Guide - Tell a Data Story** - Activity Guide [Make a Copy](#)

# Teaching Guide

## Warm Up (0 mins)

**Do This:** Move to the activity portion of the lesson.

## Activity (35 mins)

**Do This:** Students work on Page 2 of the Project Guide that covers the "New Information" step of the Data Analysis Process. Encourage students to be thorough in their answers. If they are struggling to answer the prompts, they may want to reconsider their visualization.

## Wrap Up (10 mins)

**Review:** Review the rubric

**Do This:** Collected completed Project Guides from students.

## Assessment: Project

Assess using the rubric on the final page of the Project Guide.

## Standards Alignment

CSTA K-12 Computer Science Standards (2017)

► **DA** - Data & Analysis



This curriculum is available under a  
Creative Commons License (CC BY-NC-SA 4.0).

If you are interested in licensing Code.org materials for commercial purposes, **contact us**.

# Lesson 9: Assessment Day

## Overview

Students complete a multiple choice assessment which covers the unit topics.

## Agenda

**Assessment (25 mins)**

Topic Coverage

**Assessment Review (20 mins)**

[View on Code Studio](#)

## Preparation

Preview the assessment questions

# Teaching Guide

## Assessment (25 mins)

📄🔒 Administer the Unit 9 Assessment, found on Code Studio. Make sure to unlock the assessment following instructions [here](#)

## Assessment Review (20 mins)

Review the answers to the assessment with the class. Discuss any questions that come up and take note of topics where students may need extra review.

### 💡 Teaching Tip

#### Topic Coverage

The College Board has provided a bank of questions to help formatively assess student understanding of the content in the framework. These questions are mapped to topics with each topic having a handful of questions available.

The College Board has a few strict guidelines about how topic questions can be used. In particular, students may not receive a grade based on performance on topic questions nor can they be used for teacher evaluation. Beyond these requirements, however, they are primarily intended to formatively assess student progress and learning as they prepare for the end of course exam.

Within our own course we recommend that you use them in a variety of ways:

- Throughout the unit assign topic questions to students related to the topics students are learning about that day or that week
- Prior to the unit assessment assign topic questions to help students practice and prepare for the summative assessment
- After the unit assessment use these topic questions to help students track their progress towards preparation for the AP assessment

Unit 9: Data	Topic 2.3 Extracting Information from Data
	Topic 2.4 Using Programs with Data
	Topic 5.3 Computing Bias
	Topic 5.4 Crowdsourcing

Click for more info: [Code.org CSP Topic Coverage](#)



This curriculum is available under a Creative Commons License (CC BY-NC-SA 4.0).

If you are interested in licensing Code.org materials for commercial purposes, **contact us**.