

# Post AP

In this unit students develop skills interpreting visual data and using spreadsheet and visualization tools to create their own digital artifacts. Through an ongoing project - the “class data tracker” - students learn how to collect and clean data, and to use a few common tools for computing aggregations and creating visualizations.

**Note:** The lessons in Chapter 2 of this unit do not have lesson plans and are only available to students on Code Studio. These lessons teach students to use the database tools included in App Lab. These lessons were originally intended to be a part of the core curriculum pathway but eventually were removed from Unit 5 due to timing concerns. As a result lesson plans were never developed. Even so many teachers have successfully used Chapter 2 of this unit in their classrooms. For help and support we highly recommend you speak with other teachers on the Code.org forum at [forum.code.org/c/csp](https://forum.code.org/c/csp)!

## Chapter 1: Manipulating and Visualizing Data

### Big Questions

- What is the relationship between data, information and knowledge?
- What are the best ways to find, see, and extract meaningful trends and patterns from raw data?
- Where and how does human bias affect the collection, processing and interpretation of data?

### Enduring Understandings

- 1.3 Computing can extend traditional forms of human expression and experience.
- 3.1 People use computer programs to process information to gain insight and knowledge.
- 3.2 Computing facilitates exploration and the discovery of connections in information.
- 3.3 There are trade offs when representing information as digital data.
- 7.1 Computing enhances communication, interaction, and cognition.
- 7.3 Computing has a global affect -- both beneficial and harmful -- on people and society.

### Week 1

#### Lesson 1: Introduction to Data

**Unplugged | External Tools | Individual and Group Discovery**

Students examine sources of data in the world around them how that data is collected. The Class Data Tracker project is introduced, and students predict what they will find after all the data has been collected.



## Lesson 2: Good and Bad Data Visualizations

Analyzing Artifacts | Group Discovery | Class Discussion

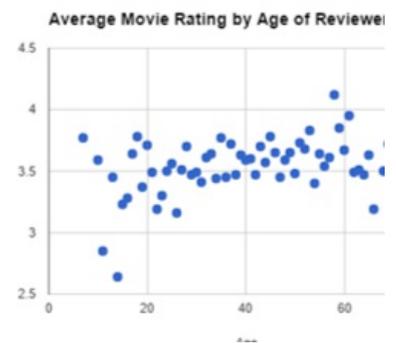
As a precursor to creating their own data visualizations, students examine collections of (mostly bad) data visualizations, rate them and discuss the characteristics of good v. bad visualizations.



## Lesson 3: Making Data Visualizations

External Tools | Individual Skill Building | Tutorial

Students follow a guide to learn how to make scatter, bar, and line charts out of provided data using a spreadsheet tool (such as Google sheets or MS Excel).

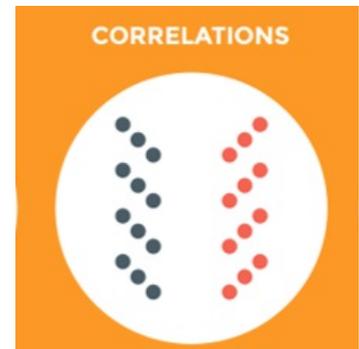


## Week 2

## Lesson 4: Discover a Data Story

External Tools | Collaborative Artifact Creation | Writing

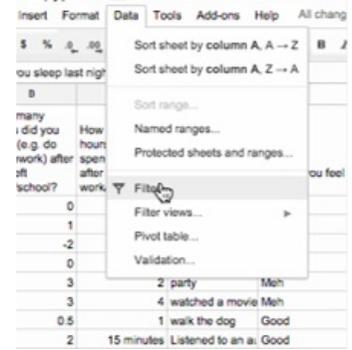
Students collaboratively investigate some datasets (provided) to “discover a data story.” Students choose one dataset, create a visualization, identify a trend, and accurately write about it.



# Lesson 5: Cleaning Data

External Tools | Analyzing | Group Skill Building

Students begin working with the data that they have been collecting for the Class Data Tracker project by first "cleaning" it to prepare it for visualization and other analyses. Each team makes their own copy of the data to examine, correct errors, categorize ambiguous items, and perform other cleaning tasks.



# Lesson 6: Creating Summary Tables

External Tools | Artifact Creation | Analyzing

Students learn how create summary tables (also known as pivot tables) from some raw datasets provided in a spreadsheet tool. Then students create and use summary tables to investigate data they've collected as a class.

Movie	Avg. Rating
Air Force One	3.615384615
Chasing Amy	4
Contact	4.05
...	...
Toy Story	3.615384615
Twelve Monkey	4
Willy Wonka	3.818181818

## Week 3

# Lesson 7: Practice PT - Tell a Data Story

Practice PT | External Tools | Artifact Creation | Project

Students continue to analyze their class tracker project data to discover, visualize, write about and present a trend or pattern they find. The writing prompts are reflective of prompts from the AP Explore Performance Task.



# Chapter Commentary

## Post AP - What's the story?

The story of this chapter is about how data can be manipulated to extract or reveal new information. Up to this point we have been focused primarily on bits and what they can be used to represent. Now we're taking a big step back to do the inverse: we want to use tools meant for viewing, manipulating, and visualizing data in order to **extract** or find new information.

The lessons in this chapter often have two things going on at once. In the background, the class is daily collecting some data about themselves (the "Class Data Tracker project") in order to accumulate data to process later on. In the interim,

students are learning about and developing skills with spreadsheet and visualization tools. The goal is for students to learn a few basic skills, see lots of examples, and then apply what they know to the **Tell A Data Story** project at the end of the chapter.

A big part of the story here is for students to understand the **computer scientist's role** in working with data, which means emphasizing how to use tools to manipulate, compute, and visualize the data. We look at things like making sure that data type choices support the way we intend to process it later (e.g. don't collect text when you need a number). Data inevitably gets "dirty" during collection and needs to be cleaned. Computers are really useful for doing some aggregations and visualizations to look for patterns. Along the way, we need to understand how human bias can be introduced at each step so that we can accurately convey what patterns in the data are or are not telling us. These activities help build toward the enduring understanding that there are trade offs when representing information as digital data.

## Our Approach to the Content

The lessons in this chapter lean heavily on external tools, especially **spreadsheets**. The benefit is that students will gain experience with **real tools and real data** for the first time. The pitfall is that, because the tools are external, they are not scaffolded or designed for learning. We have tried to provide tutorials and curated data sets to ease the burden as much as possible, but ultimately you're operating in the real world. While confined to the world of your classroom, the Class Data Tracker project should provide some **authentic** examples, scenarios, and sometimes headaches related to data collection and processing in the real world.

As the teacher it's important to keep in mind the goals of CS Principles because it can be enticing with these lessons to dig into "hardcore" data analysis techniques and statistics. While these are important, they are beyond the scope of CS Principles. Thus, we treat data analysis and statistics a bit like an electric fence: get close, but don't touch. Students should be able to extract interesting things as the result of letting the tools do the work. We provide some large sets of curated data that came from real sources. The data is big enough that you have to apply some computation to make sense of it. We show how to use spreadsheets to do basic aggregations (such as grouping, counting, clustering) and computations (such as average, median, etc.), without turning it into a lesson on statistics and data analysis. We want to build toward the enduring understanding that computing facilitates exploration and the discovery of connections in information.

The idea behind the **Class Data Tracker** project is that we have found that when students work with data that they collected themselves it is easier and **intrinsically motivating** for students to dig in. To accumulate enough data, we collect it in increments during the time they're building up other skills with data tools. You should connect the skills students are learning in the exercises to similar things they might do with the class tracker data for the **Tell a Data Story** project.



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: Introduction to Data

Unplugged | External Tools | Individual and Group Discovery

## Overview

In this kickoff to the Data Unit, students begin thinking about how data is collected and what can be learned from it. To begin the lesson, students will take a short online quiz that supposedly determines something interesting or funny about their personality. Afterwards they will brainstorm other sources of data in the world around them, leading to a discussion of how that data is collected. This discussion motivates the introduction of the Class Data Tracker project that will run through the second half of this unit. Students will take the survey for the first time and be shown what the results will look like. To close the class, students will make predictions of what they will find when all the data has been collected in a couple weeks.

## Purpose

This lesson introduces many of the lessons and themes that will run through the unit. Students are introduced to the Class Data Tracker and the fact that they will be collecting and analyzing their own data in a couple weeks. They also begin thinking about the many ways data impacts their lives and how it can be used. While the primary goal of this lesson is to get ideas and processes in place for the rest of the unit, there are many places where students can start asking interesting questions about where and how data is collected, who is collecting it, and how they are using it.

## Agenda

### Teacher Setup (10 mins)

Teacher Setup Guide for Data Tracker Project

### Getting Started (20 mins)

Pop “Quiz”

### Activity (25 mins)

Introduce Class Data Tracker survey.

### Wrap-up

[View on Code Studio](#)

## Objectives

Students will be able to:

- Develop a hypothesis about student behavior over time, based on a small sample of data.
- Describe sources of data appropriate for performing computations.

## Preparation

Review the **Data Tools Resources** for this lesson (including Excel support)

Teacher Setup for Google Forms (see Teacher Setup in Teaching Guide)

## Links

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

**For the Teacher**

- **Class Data Tracker Setup Guide** - Google Form Setup (includes template)

[Make a Copy](#)

**For the Students**

- **How Much of a Left and Right-Brained Person Are You?** - [Link](#)

# Teaching Guide

## Teacher Setup (10 mins)

### Teacher Setup Guide for Data Tracker Project

This lesson requires a one-time special setup in order to create a form for data collection with the students in your class. Once you have it setup you will use it for several weeks.

 **Please use: *Class Data Tracker Setup Guide - Google Form Setup (includes template)***

Please click the link to see the full **Class Data Tracker Setup Guide - Google Form Setup (includes template)** .

In a nutshell the guide has you:

1. Make a copy of a Google Form (**short link to template**)
2. Share a link to the form with your students

There are also **notes on editing the survey questions** if you want to -- we chose the questions so that certain properties would emerge later on. If you want to change the questions just ensure that you'll get the same properties or later lessons might not work.

After the setup you should have:

- A copy of a Google form in your Google Drive
- A spreadsheet that will collect responses from the form in your Google Drive
- A link that students should use to fill out the survey

Students fill out this form **every day** or as frequently as possible over the next few weeks. We will look at this later in the unit. You can place the form and spreadsheet documents wherever you like in your Google Drive. They are yours now.

## Getting Started (20 mins)

### **Opening Remarks**

**Transitioning to a new chapter.** These remarks are meant to help you make a bridge between the **Encode and Experience** activity and this chapter about manipulating and visualizing data.

- The last project you did (Encoding an Experience) was about organizing and structuring digital data to represent complex information.
- You did it by thinking about bits.
- In reality we typically don't have to break digital data down all the way to bits in order to work with it, but understanding that digital data at its root is just bits **gives you insights** into working with larger data sets.
- We are about to embark on a new series of lessons where you will work with real data sets and learn how to use tools to explore and extract information and knowledge from the data.
- One way we think about it is learning how **tell stories with data**. We start today!

### **Pop “Quiz”**

Before saying anything, point students to this online quiz and have them complete it: **How Much of a Left and Right-Brained Person Are You? - Link**

**Share Results:** Allow students to share and compare their percentages of left and right-brainedness. It should be mildly amusing. The point of this little exercise will be revealed after the discussion.

### Discussion

Get students to start thinking about where they interact with and produce data in their lives, by looking at their past experiences with online quizzes and surveys, to bridge the gap to a long-term class data collection activity.

### **Remarks**

This unit will address the topic of data more deeply. In computing, we're interested in where data comes from, what structure or formats it comes in, and most importantly, what kind of knowledge or information we can extract from that data using computational tools.

### Prompt:

**"People say there is data all around us. What do you think that means? Brainstorm as many examples of data as you can think of."**

For each one, try to answer:

- Who is generating the data?
- Where is the data being stored or saved? Who owns it?

### Discussion

- Give students 2 minutes to jot down ideas before sharing with a neighbor.
- Do a whip-around to get ideas out in the air, perhaps writing them on the board.
- Student responses will vary widely and may be related to:
  - cell phone data plans
  - science experiments
  - GPS tracking
  - online shopping data
  - taxes or accounting info
  - sports data

#### Discussion

- Make sure to point out that, for most of these examples, people are generating the data through their own actions, though sometimes they might not be aware of it.
- In most cases this data is stored somewhere else, and by someone else.
- The point to make is not necessarily a concern for privacy (yet) but simply the fact that there is lots of data gathered by individuals and organizations, which makes it possible to compute with/on.
- Some knowledge could be extracted from that data.

### Transitional Remarks

Good, you identified all kinds of places that data comes from. In this unit we'll be looking at lots of those same examples and learning a bit about how to use, manipulate and visualize data with computational tools.

In Computer Science, sometimes we can have the computer itself generate data for us. Later in the course when we get to programming, we'll write programs that generate a lot of data.

But there are other kinds of data that can't be generated by the computer. In particular, data about people and how they act in the real world is hard to capture without just asking them. So that's what a lot of tools online do. They try to capture people's responses to things because the data, in aggregate, might contain useful information that could be extracted.

That "dumb" online quiz you took at the beginning of class is an example. These quizzes ask people to reveal things about themselves, their preferences, likes and dislikes. This is data! While these online quizzes are probably innocuous, some interesting things about people could probably be discovered if the data were analyzed.

As a class, we're going to do something similar...

## Activity (25 mins)

**Setup Reminder:** Make sure you have prepared the Google form, and have the share link ready ahead of time. See notes above.

### Remarks

As our first adventure into data, each of you is going to complete a short survey. Surveys are one of the best ways to collect data from people, and they are functionally no different from an online poll, funny quiz, or anything else that asks you for your opinion. We're going to use our own survey, so that we can collect and see all the data.

#### Activity

Introduce the class data tracker project. The class will collect data about themselves so that students can see trends and patterns in the class's behavior over time.

## Introduce Class Data Tracker survey.

**Distribute:** Share the survey link with your students and have them complete it once.

**Display the Initial Responses:** Once everyone has filled out the survey, show them a glimpse of the results. You can find the results from your survey by clicking the Responses tab next to the Questions tab at the top of the form you made.

QUESTIONS

RESPONSES

Display the responses on the board. Scroll through them, giving students a chance to see the data. Try not to get hung up on issues of formatting, like a student who responded “seven hours” instead of “7” or “7 hours.”

You may want to show the raw spreadsheet view instead of, or in addition to, the default “dump” of responses shown in the form.

**Briefly Discuss:** Have students look at the results from the survey and discuss what they notice.

- What do you notice?
- What was surprising?
- What do the results tell you about you and your answers?
- What other information would you like?
- What kind of questions would we need to ask to find out more information?

**Explain:**

You are going to complete this survey every day in class for the next several weeks. By the end, we should have several hundred entries. You’ve seen the questions and have taken a quick glimpse at the results. What do you think we might be able to find out in a few weeks?

**Prompt:**

- 💡 **“Write down one or two hypotheses (predictions) about what we might be able to find out about our class, assuming that everyone fills out this survey every day for a few weeks.”**

Transition to wrap up.

#### 💡 Teaching Tip

If necessary, introduce (or review) the term hypothesis with your students. The CSP Framework has a learning objective that reads: 3.1.1 Find patterns, and test hypotheses about digitally processed information to gain insight and knowledge. [P4] We will come back to these hypotheses when we look at the data in earnest a few lessons down the line.

Many will have probably seen the word in a science class. The **Merriam-Webster Dictionary** says a hypothesis is “an idea or theory that is not proven but that leads to further study or discussion.”

## Wrap-up

**Share:**

- 🗨️ Do a quick share-out of students’ hypotheses about what the class data will show in a few weeks.

**“What kinds of predictions did you make?”**

- Student responses will focus on different aspects of the data.
- Anything related to time spent doing things outside of school and how it makes you feel is fair game.

#### 🎤 Remarks

- Those are all interesting ideas.
- Many of them will require us to perform some computations on the results to find the answers, or spot other trends or patterns.
- Over the coming weeks, we’ll collect this data, and over that time, you’ll learn some things about how to process and visualize data like this, so you can see for yourself what kinds of knowledge the data holds.

Welcome to data.

#### 💬 Discussion

Foreshadow the class data tracker project and the rest of the Unit.

In student’s hypotheses: try to focus on hypotheses that hinge on a relationship between two elements of the data. For example:

- people who get more sleep tend to feel better
- predictions about trends or other patterns (e.g., I think most people will go to the movies to relax, but only on weekends).

# Standards Alignment

## Computer Science Principles

- ▶ **3.1** - People use computer programs to process information to gain insight and knowledge.
- ▶ **3.2** - Computing facilitates exploration and the discovery of connections in information.
- ▶ **7.3** - Computing has a global affect -- both beneficial and harmful -- on people and society.

## 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 2: Good and Bad Data Visualizations

Analyzing Artifacts | Group Discovery | Class Discussion

## Overview

This is a pretty fun lesson that has two main parts. First students warm up by reflecting on the reasons data visualizations are used to communicate about data. This leads to the main activity in which students look at some collections of (mostly bad) data visualizations, rate them, explain why a good one is effective, and also suggest a fix for a bad one.

In the second part of class students compare their experiences and create a class list of common faults and best practices for creating data visualizations. Finally, students review and read the first few pages of **Data Visualization 101: How to design charts and graphs** to see some basic principles of good data visualizations and see how they compare with the list the class came up with.

## Purpose

An important skill is the ability to critically evaluate information. As our world is increasingly filled with data, more and more the information from that data is conveyed through visualizations. Visualization is useful for both **discovery** of connections and trends and also **communication** - both are potentially aspects of the Explore Performance Task. In this lesson we will focus on the communication aspects of visualization.

Interpreting data visualizations is not typically thought of as a core computer science skill, but it is certainly an important one in an age of digital data. Computing has enabled massive amounts of information to be automatically collected, aggregated, analyzed, and visualized. Visualizations are useful in helping humans understand large amounts of data quickly, and they are useful communication tools when presenting findings about a collection of data. Not all visualizations are created equal, however, and in many cases the type of visualization used may distract or even mislead the reader.

As both creators and consumers of data visualizations, students need to be on the lookout for these common pitfalls. This will allow them to be savvier readers of data visualizations, and more effective communicators when creating visualizations of their own.

## Agenda

**Getting Started (10 mins)**

**Fill out class tracker survey**

[View on Code Studio](#)

## Objectives

**Students will be able to:**

- Identify an effective data visualization and give justification.
- Collaborate to investigate and evaluate a data visualization.
- Suggest an appropriate visualization for some data.
- Evaluate a data visualization for effectiveness of communication.
- Identify a poor data visualization and give justification.

## Preparation

Print a copy of **Data Visualization Scorecard - Worksheet** for each student.

## Links

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

**For the Students**

- **Data Visualization 101: How to design charts and graphs** - Link
- **Data Visualization Scorecard** - Worksheet [Make a Copy](#)
- Data Visualization Collection A & B (in Code Studio)

**Think-Pair-Share: Why Make Visualizations?**

**Activity (25 mins)**

**Review and Rate Data Visualizations**

**Debrief: What makes a good/bad data visualization?**

**Make a table of good v. bad visualization characteristics**

**Wrap-up (15 mins)**

**Data Visualization 101 discussion**

**Assessment**

**Assessment Possibilities**

**Extended Learning**

# Teaching Guide

## Getting Started (10 mins)

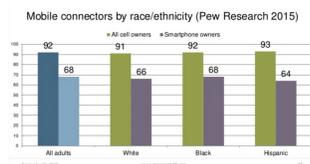
### Fill out class tracker survey

**Survey Reminder:** Give students a few minutes to fill out the class tracker survey that you started in **Lesson 7 - Introduction to Data**.

### Think-Pair-Share: Why Make Visualizations?

#### Remarks

Yesterday you looked at a bunch of data from the Pew Research Center that was all presented visually in graphs and charts. The question is: why? Why did they choose to make a bunch of charts and graphs rather than just showing the raw data itself?



#### Prompts:

Here are two related prompts to respond to:

- "Why did Pew Research choose to make a bunch of charts and graphs rather than just showing the raw data itself?"
- "List a few advantages and disadvantages (at least 2 for each) of using visualizations to communicate data"

**Pair:** Have students share with an elbow part.

**Share:** Draw out responses from the a whole class to generate common themes.

 Here are a few notes to help guide the discussion:

#### Why visualize?"

- Student responses should focus on **communication**. You choose to make a chart because you think it's a better or more effective way to communicate information than using raw data.

#### Advantages / Disadvantages

- Advantages: pictures allow you to compare things more easily, easier to see trends or patterns, can focus on, or highlight, particular aspects of the data that are important
- Disadvantages: easy to mislead or miscommunicate, removes details that might be important or valuable, sometimes very dense - takes a while to study to understand what it means.

#### Discussion

The goal here is to do a quick thinking prompt to activate prior knowledge about visualizations. There are two key points to draw out:

- Visualization is about **communication**. The goal of any data visualization is to transform data into useful **information**.
- There are **both** advantages and disadvantages to data visualizations - some of the disadvantages might not be as clear at this point but they will be after this lesson. The disadvantages mostly stem from the fact that human error and bias can be introduced when trying to communicate.

## Activity (25 mins)

#### Remarks

Making even a small visualization may have been surprisingly challenging and varied.

In fact, even experienced data analysts can end up obscuring their message when they make data visualizations.

To better understand some of the skills we just read about, we are going to evaluate a collection of data visualizations to determine how well they communicate their message.

## Review and Rate Data Visualizations

**Pair:** Partner students who will work through the worksheet together.

**Assign:** There are two different collections of data visualizations. Each pair of students should be assigned to evaluate one of **either**:

- Data Visualization **Collection A**

or

- Data Visualization **Collection B**

**Links to the separate collections can be found in Code Studio**

**Distribute:** Worksheet - **Data Visualization Scorecard - Worksheet**

(There is a link to the worksheet in Code Studio, but this is one you probably want printed out.)



**Transition to Good and Bad Visualizations on Code Studio**

The worksheet asks pairs of students to **collaborate in reviewing the data visualizations** :

- Give a rating from “Great” to “Horrible” for 15 different data visualizations.
- Choose the best (or favorite one) and explain why it effectively tells a story or communicates some data.
- Choose the worst (or a bad one) and explain why it’s ineffective.
  - **Optional:** Suggest (via a sketch) a better visual that could represent the same data.

**Share:**

- 🔗 After completing the worksheet, have each group share the best and worst image from their set with another group. Groups should focus on how they would fix the worst visualization they chose. Share and exchange ideas about different ways to visualize the data.

### 💡 Teaching Tip

Compare with Different Groups: Collection A and Collection B have different sets of visualizations. There are benefits to discussing with both a group that used the same collection of visualizations and with a group that used a different one. Time allowing, encourage groups to share their findings with groups who used both collections of visualizations.

## Debrief: What makes a good/bad data visualization?

Have a discussion with two main parts...

**Part one: Share out best/worst**

- Ask student pairs to share the graphic they rated the worst and best.
- Focus on one or two and have all students look at it (bring up on a projector or have all students bring it up on screen).
- Ask students to justify or give reasons why they rated the graphic highly or poorly.

### 💡 Teaching Tip

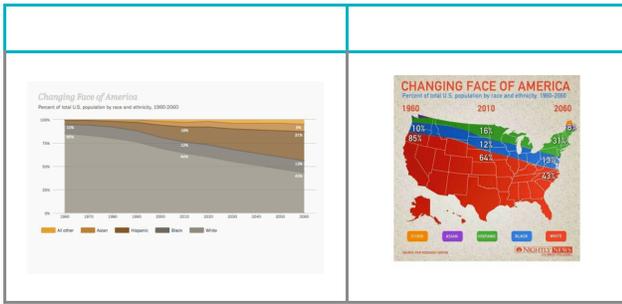
To be creative with the share-out, you could:

- Have students vote publicly on each one in some way
- Have the groups that looked at Collection A all get together and figure out a best/worst to show to the other group(s) that looked at Collection B, and vice versa.

**Part 2: Discuss graphic number 5 - The Changing Face of America**

**Graphic number 5 in both collections** are two different displays of the same data. For reference, here are snapshots of both graphics.

You might **display these** on the screen, **or have students look together** by sitting next to each other and opening up the graphic from each collection.



**Prompt:**

**"Can everyone please look at graphic number 5 in their collection. It's called "The Changing Face of America."**

**Whip around:**

- How did different groups rate this graphic?
- What data is presented?
- What is the difference between the two visualizations?

**Students should recognize that the two graphics are attempting to represent the same data.**

## Make a table of good v. bad visualization characteristics

Prompts:

- Following the principles of good data visualization, which one would we say is better?
- What makes the good one good and the bad one bad?

As students respond, steer the discussion toward generating general characteristics of good and bad visualizations. Make a simple chart that everyone can see.

Something like this...

Good	Bad
<ul style="list-style-type: none"> <li>• simple</li> <li>• easy to read</li> <li>• a basic graph that makes a simple point</li> <li>• ...etc...</li> </ul>	<ul style="list-style-type: none"> <li>• complicated</li> <li>• confusing colors</li> <li>• too much text</li> <li>• ...etc...</li> </ul>

## Wrap-up (15 mins)

### Data Visualization 101 discussion

#### Remarks

We're going to be making some of our own visualizations of data very soon. To help us do that, we're going to look at some helpful tips for effectively communicating with data visualization.

#### Discussion

Summarize the findings from the visualizations work by examining two visualizations that present the same information in starkly different ways. As part of the discussion...

1. Students should recognize that the two graphics are plotting the same data.
2. Build a list of good/bad properties of data visualizations.

#### Teaching Tip

Strategies for bringing out characteristics of good/bad visualizations:

- fill in a chart at the front of the room as students talk
- Have students write ideas on post-its and attach to a centrally located chart as basis of discussion
- Have students add ideas and comments to a shared online document

#### Discussion

Establish some rules of thumb for visual displays of data. Students should recognize that some types of charts are more appropriate than others, depending on the nature of the data or the message the author is trying to convey.

💡 **Distribute: Data Visualization 101: How to design charts and graphs - Link.** Students should read the first 4 pages of this document.

💬 **Discuss:** What are the key take-aways from this guide?

Some key ideas that should come up:

- Choosing the right way to visualize data is essential to communicating your ideas.
- There are stories in data; visualization helps you tell them.
- Before understanding visualizations, you must understand the types of data that can be visualized and their relationships to each other.
- Certain chart types are right for certain situations, depending on the data.

### 🗣️ **Remarks**

The Data Visualization 101 guide is a resource for you (students).

The rest of the guide goes into some specifics of different chart types.

You should keep this guide at your side as you review visualizations data, and when you develop your own in the future.

**Further Discussion Points:** What else did we learn about data visualization today?

- What are the benefits of visualizing data?
- Can we characterize common mistakes in visualizations to which we gave low ratings?
- Can we characterize common strengths in effective visualizations?
- Not all visualizations were charts; what other types are there?
- As you embark on making your own visualization, what do you want to keep in mind so that you can avoid rookie mistakes?

### 💡 Teaching Tip

**Reading Strategies:** Students can read individually, in partners, or as a whole class. The guide is not particularly long, but you'll want all students to have had a chance to look through those pages before the discussion. Let students know ahead of time that you'll be discussing the reading and ask them to pick one or two key points as they are going through.

## Assessment

### Assessment Possibilities

- Assessment Idea: show students a visualization and have them analyze it, using the table of characteristics of good/bad visualizations to justify their opinion.

#### Performance Task-style reflection question

**Choose the visualization that you thought was the best or worst (pick one) from the ones you saw in class and do the following:**

- **Describe the visualization so the reader knows which one you are talking about (example: "Collection A #2 -- Average divorce rates in America")**
- **Say whether this was the best or worst visualization for you and why. Justify your opinion by citing principles of visualizations that you have learned about. Use the visualization 101 guide as a resource.**
- **Try to keep your response to around 100 words (about 3-5 sentences).**

## Extended Learning

If you want additional sources of data visualizations, consider the following sources:

- Daily Infographic: <http://www.dailyinfographic.com/>
- Infographics Archive: <http://www.infographicsarchive.com/>

## Standards Alignment

## Computer Science Principles

- ▶ **1.2** - Computing enables people to use creative development processes to create computational artifacts for creative expression or to solve a problem.
- ▶ **3.1** - People use computer programs to process information to gain insight and knowledge.

## CSTA K-12 Computer Science Standards (2017)

- ▶ **DA** - Data & Analysis
- ▶ **IC** - Impacts of Computing



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: Making Data Visualizations

External Tools | Individual Skill Building | Tutorial

## Overview

Now that students have had the chance to see and evaluate various data visualizations, they will learn to make visualizations of their own. This lesson teaches students how to build visualizations from provided datasets. The levels in Code Studio provide a detailed walkthrough of how to use Google Sheets to create several different kinds of charts. While this lesson focuses on the Google Sheets tool, other tools may be substituted at the teacher's discretion, and MS Excel support is coming soon to the lesson.

The main activity teaches students to build different chart types (scatter, line, and bar charts) from a single data set. It should be emphasized to students that the purpose of this lesson is to explore and experiment with creating different types of visualizations, not to build the perfect chart. Students will have a chance to create and customize their own charts. At the end of class, students compare their custom visualizations with those of their classmates.

## Purpose

Being able to create meaningful data visualizations is extremely important in order to effectively communicate information about large data sets. It's also important to be able to use visualizations to simply "look" at data that is too complex to make sense of by looking at the raw data alone. Any computer scientist working with data should have some skills and facility with producing visualizations of the data to get a sense of what it contains. Visualizing the data allows you to see patterns, trends or relationships you might otherwise not.

The most important piece of this lesson is not learning to create the prettiest chart; it's about using charts to "tell the story" of what's really going on in the data. Different charts are more or less appropriate for communicating this story, depending on the data. The point of having students explore different chart types is to help them build visualizations that reveal trends or connections in the data that are too hard to see by just looking at a data table in a spreadsheet.

## Agenda

### Getting Started

#### Survey Reminder

#### Using visualization to discover connections and patterns

#### Make a Quick Visualization

### Activity

[View on Code Studio](#)

## Objectives

Students will be able to:

- Select the appropriate type of data visualization to discover trends and patterns within a dataset.
- Create a bar, line, and scatter chart from a dataset using a computational tool.
- Use the settings of a data visualization tool to manipulate and refine the features of a data visualization.

## Preparation

[Data Tools Resources](#) (including Excel support)

## Links

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

### For the Teacher

- **KEY - Making Data Visualizations Target Charts - 2018** - Answer Key

### For the Students

- **Data Visualization 101: How to design charts and graphs** - Link
- **MovieRating\_avgRatingByAgeByGender.csv** - Data Set ([download](#))

**Make scatter, line, bar, and custom charts**

**Wrap-up (10 mins)**

**Compare with a partner**

**Assessment**

**Assessment Possibilities**

**Extended Learning**

# Teaching Guide

## Getting Started

### Survey Reminder

**Survey Reminder:** Give students a few minutes to fill out the class tracker survey that you started in **Lesson 7 - Introduction to Data**.

### Using visualization to discover connections and patterns

**Prompt:**

**"Do you have to use a computer to create a data visualization? What are some reasons that you need to use a computer to manipulate data?"**

#### Briefly Share and discuss responses

- Do a quick think-pair-share (or other strategy)

#### *Transitional Remark*

Taking data from its raw state to the point where you can create a meaningful visualization involves several steps. Today we're going to use visualization in attempt to discover things in the data we might not otherwise see.

It takes practice to create good visualizations. Today, we'll get our feet wet by learning to create charts using Google Sheets.

### Make a Quick Visualization

#### *Remarks*

When trying to understand data, having a visualization, or picture of it, is often much more effective at communicating information than the raw data itself.

Making a good visualization of data is often challenging but can be fun and very creative, and we're about to start making our own. Let's try one, quickly.

#### **Scenario:**

Here is some data: **On some survey 2,000 people were asked, "What do you do when you're bored?". Here are the most common responses by age group.**

age	most common response	number	out of
18 and under	texting	157	500
19-64	watching TV	247	1200

#### Discussion

We want to motivate students' desire to create some visualizations on their own. Build on the "Good/Bad Visualizations" lesson. Some responses students might give:

- A large data set is too big to understand by looking at a table in a spreadsheet.
- Creating a data visualization with a computer is faster and more accurate than creating one by hand.

#### warm up

This is intended to only be a brief activity to illuminate issues around making visualizations and how much variety there can be, and letting the students be creative and share with each other.

#### Teaching Tip

Please note: this data is completely fabricated and is only intended to serve the purposes of the warm up. It is intentionally slightly ambiguous. If students ask questions seeking clarification that's a good sign, but you might have to simply respond: "Well, this is the data we have".

There are no right or wrong answers here as long as students attempt to represent the data in a different way **somehow**.

age	most common response	number	out of
65+	reading	54	300
all ages	talking with friends	451	2000

For example: of the 1200 people surveyed between the ages of 19-64, 247 said "watching TV" which was the most common of any other responses to the question for that age group.

**Prompt:**

- "Take a few minutes by yourself and try to make a visual, graphical, explanation of this data. Try to communicate something about through drawing while remaining true to the results of the data."
- Give students 3-5 minutes to draw.

**Compare and Discuss:**

- Have students compare what they drew with an elbow partner and point out similarities and differences.

Prompts:

- "In this exercise what was challenging?"
- "What kinds of things were visually effective at communicating information?" <-- ALT: "What were the characteristics of the visualizations that effectively communicated this information visually?"

## Activity

### Make scatter, line, bar, and custom charts

 **Transition to Making Data Visualizations on Code Studio**

- The "Activity Guide" for this lesson is all laid out in Code Studio.
- Put students into pairs and send them to Code Studio.
- The steps students go through are laid out below.
- Please note the purpose and teaching tips on this lesson for perspective.

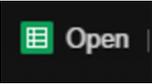
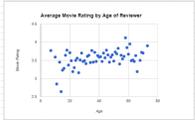
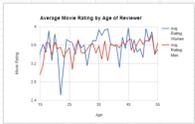
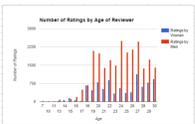
 **Teaching Tip**

Remember the point here is not to make the prettiest chart, but choose the chart type that makes the most sense for the data you've got and the story you're trying to tell.

You can also point out to students that finding "no correlation" or "no relationship" is actually just as interesting as finding a strong correlation or relationship. For example, if you examine the difference between men and women in average rating of Star Wars, you will see virtually no difference! That's interesting!

While students are working, circulate the room to help and encourage.

- Use **KEY - Making Data Visualizations Target Charts - Answer Key** which shows what students should be trying to create. It also explains a few common mistakes that students might make.

Teacher Code Studio Reference	
<p>Students are asked to make a copy of the data set in their Google Drive. (Students must be logged into Google Drive for this step to work.) When they open the link to the CSV file, they can click the “Open” button next to the green Google Sheets logo, which will make a copy of the CSV in their personal Drive folder.</p>	
<p>Students follow step-by-step instructions to create a scatter plot showing the average movie rating by age of reviewer</p>	
<p>Students follow step-by-step instructions to create a line chart showing the average movie rating by age of reviewer, broken down by gender.</p>	
<p>Students follow step-by-step instructions to create a bar chart showing the number of ratings by age of reviewer, broken down by gender.</p>	
<p>Students experiment with creating their own charts on the same data set  <b>NOTE:</b> they'll get a chance to explore many different data sets in the next lesson. It should be emphasized that the purpose of this part of the lesson is to freely explore the chart tool and discover connections in the data; students should not fixate on creating the perfect chart.</p>	

## Wrap-up (10 mins)

### Compare with a partner

With partners or in small groups, have students discuss the following prompt. Once students have shared with each other, have students report back to the class about the charts they made and what they learned.

**Prompt:** What was the most interesting visualization you were able to create? What did it help you discover about the data?

## Assessment

### Assessment Possibilities

- Score or review a written response to the reflection prompt from the wrap up (also found in code studio)
- Make a simple rubric (a checklist basically) for the steps of the activity that students were supposed to go through:
  - Scatter Plot
  - Line Chart
  - Bar Chart
  - Optional: something on their own

# Extended Learning

If you want additional sources of data visualizations, consider the following sources:

- Daily Infographic: <http://www.dailyinfographic.com/>
- Infographics Archive: <http://www.infographicsarchive.com/>

## Standards Alignment

### Computer Science Principles

- ▶ **1.2** - Computing enables people to use creative development processes to create computational artifacts for creative expression or to solve a problem.
- ▶ **3.1** - People use computer programs to process information to gain insight and knowledge.

### CSTA K-12 Computer Science Standards (2017)

- ▶ **DA** - Data & Analysis
- ▶ **IC** - Impacts of Computing



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: Discover a Data Story

External Tools | Collaborative Artifact Creation | Writing

## Overview

In this lesson, students will collaboratively investigate some datasets and use visualization tools to “discover a data story.” The lesson assumes that students know how to use some kind of visualization tool - in the previous lesson we used the charting tools of a basic spreadsheet program. Students should be working with a partner but without much teacher hand-holding. Most of the time should be spent with students poking around the data and trying to discover connections and trends using data visualization tools. It is up to them to discover a trend, make a chart, and accurately write about it.

## Purpose

Being able to look at large sets of data and use visualization as a tool for discovery is a common task that many people who work with data do on a daily basis. A computer scientist should have decent facility with using tools opening and browsing large datasets, and doing some cursory exploration to see what’s there. The computer scientist should be familiar enough with the tools to, over time, develop some instincts about data, how it’s collected, the kinds of formats it comes in, and how that affects what can or cannot be done to visualize it.

## Agenda

### Getting Started (10 mins)

Fill out class tracker survey  
Visualization as a discovery tool  
Quick Investigation of a sample dataset

### Activity (40 mins)

Discover a Data Story

### Wrap Up (10 mins)

Share your data stories

### Assessment

Assessment Possibilities

[View on Code Studio](#)

## Objectives

Students will be able to:

- Collaboratively investigate a dataset.
- Create a visualization (chart) from provided data.
- Identify possible trends or connections in a data set by creating visualizations of it.
- Accurately communicate about a visualization of their own creation.

## Preparation

Print a copy of **Activity Guide -**

**Discover a Data Story - Activity Guide** for each student (or provide link)

Print a copy of **Data Visualization 101:**

**How to design charts and graphs - Link** for each student (or provide link) NOTE: this may have already been printed for lesson 10

## Links

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

**For the Students**

- **Activity Guide - Discover a Data Story - Activity Guide** [Make a Copy](#)
- **Data Visualization 101: How to design charts and graphs - Link**
- **Rubric - Discover a Data Story - Rubric** [Make a Copy](#)
- **Data sets folder - Folder**

# Teaching Guide

## Getting Started (10 mins)

### Fill out class tracker survey

**Survey Reminder:** Give students a few minutes to fill out the class tracker survey that you started in **Lesson 7 - Introduction to Data**.

### Visualization as a discovery tool

#### **Remarks**

In the previous lesson, we learned how to use a data visualization tool to create a visualization. Sometimes data in its raw state is simply too big to be able to look at and derive any meaning. Even when the data is summarized in a table, it can be difficult to “see” what the data shows.

Today we’re going to see how visualizing data can be a useful tool for discovery. In today’s activity, you and a partner will investigate some sets of data on your own and use visualization to discover a connection or trend.

### Quick Investigation of a sample dataset

For today’s work there are several datasets for you to choose from.

- We’re going to **take 5 minutes** to poke around in one of the datasets to see how it’s structured.
- Then we’ll come back together to get some terms straight before discovering further.

 Go to Code Studio

1. Find the link to the **“Personality” dataset** and open the folder.
2. Find and open the **README file**.
3. Find and open the **rawData.csv file**.
4. Find and open **one other .csv file** - there are a few.

**Discuss:** What’s in the folder for a dataset?

After students have had a few minutes to poke around, make sure the group understands what these files are.

You can use a think-pair-share or a simple whole group discussion to get the details out.

Ask the questions below, explanations are provided for you

#### **“What’s the README file?”**

- Most datasets, when you download them, contain a README file.
- The README file is just a plain text document that gives some background information about the dataset, how it was collected, and what the column headings mean.
- The README is a good first stop when trying to understand exactly what a dataset contains.

#### **“What’s the rawData.csv file?”**

- For the datasets we provide, each folder contains a “raw” dataset, which is the original data, as it was collected.
- Recall that **.csv** stands for “comma-separated values.” CSV is a common, plain text format for distributing datasets.

#### **“What’s in the other CSV files?”**

- The other files are what we call **summary tables**.
- These are tables that were created by running some computations on the raw data to do things like count, average, sum, compare, and categorize the data in interesting ways.
- It is likely that these summary tables will be the data you use to create your visualizations.

## Activity (40 mins)

## Discover a Data Story

**Pair:** Put students in pairs or small groups to explore the datasets

### Remarks

With your partner explore the datasets and choose one you'd like to learn more about. Make sure you

- Read the README to understand the raw data that was collected
- Look at the summary tables provided for your dataset.
- Repeat these steps with additional datasets
- Choose one to explore more deeply.

### Discover a Data Story

**Distribute:** Activity Guide - **Activity Guide - Discover a Data Story - Activity Guide**

There is a link to this guide in Code Studio.

You may choose to have students make their own digital copies of this document and work on it there as well.

The activity guide asks students to:

- Pick a dataset
- Use visualization tools to “discover a data story”
- Prepare one (or two) to present
- Respond to prompts

## Wrap Up (10 mins)

### Share your data stories

Have students share their data stories with each other or with the whole class. A pair should:

- Show the visualization they made.
- Explain what it shows.
- Explain the possible story it tells.

## Assessment

### Assessment Possibilities

**Use the rubric to score the activity guide**

**You may choose** to collect the second page of the

Activity Guide and score it using the **Rubric - Discover a Data Story - Rubric** provided.

**Note:** Collecting and scoring the Activity Guide is optional.

- The intent of this activity is NOT to make a huge project out of it.
- The goal is simply to come away with some artifact that you might assess.

### Teaching Tip

#### A Note on distributing the Activity Guide

The first section of the activity guide contains the instructions above. It's suggested that students start exploring the datasets **before you distribute the activity guide** so they don't lose momentum.

You might choose to **assign the datasets to groups**. This cuts down on student choice, but might save time if students are taking a while to settle on which dataset they want to use.

#### While students are working:

- Remind students of the existence of the guide: **Data Visualization 101: How to design charts and graphs - Link**.
- Most of the students' time should be spent on working collaboratively to visualize data in different ways.
- Encourage and remind students that an “interesting” finding doesn't necessarily mean finding something world-changing or mind-blowing. The data is so big and hard to “see” that **simply making a clear chart that gives some kind of view into the data is interesting**

### Teaching Tip

For student sharing, there are a number of different things you could do, depending on your needs and classroom dynamic. Here are a few suggestions.

- Have groups that used the same dataset share with each other.
- Have each group share with one or two groups who used a different dataset.
- Highlight one or two pairs' work by asking them to present to the whole class.

- It might be sufficient for students to share what they created in class rather than submitting the worksheet.

### **Personal Reflection: Collaboration**

**This prompt is also provided on Code Studio**

(NOTE: The following is modification of one of the prompts given on the AP Create Performance task.)

**Prompt: Describe the development process of discovering your data story and creating a visualization. Describe the difficulties and/or opportunities you encountered along the way, and describe the collaborative process between you and your partner.**

**Please limit your response to about 200 words.**

## **Standards Alignment**

### **Computer Science Principles**

- ▶ **1.1** - Creative development can be an essential process for creating computational artifacts.
- ▶ **1.2** - Computing enables people to use creative development processes to create computational artifacts for creative expression or to solve a problem.
- ▶ **1.3** - Computing can extend traditional forms of human expression and experience.
- ▶ **3.1** - People use computer programs to process information to gain insight and knowledge.

### **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 5: Cleaning Data

External Tools | Analyzing | Group Skill Building

## Overview

In this lesson, students begin working with the data that they have been collecting since the first lesson of the chapter in the class "data tracker." They are introduced to the first step in analyzing data: cleaning the data. Students will follow a guide in Code Studio, which demonstrates the common techniques of filtering and sorting data to familiarize themselves with its contents. Then they will correct errors they find in the data by either hand-correcting invalid values or deleting them. Finally they will categorize any free-text columns that were collected to prepare them for analysis. This lesson introduces many new skills with spreadsheets and reveals the sometimes subjective nature of data analysis.

## Purpose

The main purpose here is have students independently apply some of the data manipulation skills (in spreadsheets) that they've learned over the past few lessons, to a new dataset that is relatively uncurated. This is the beginning of the process of "extracting knowledge from data": look at the data and clean it up so that you can process it using computational tools.

Using computational tools to analyze data has made it much easier to find trends and patterns in large datasets. When preparing data for this kind of analysis, however, it's important to remember that the computer is much less "intelligent" than we might imagine. Small discrepancies in the data may prevent accurate interpretation of trends and patterns or can even make it impossible to use the data in computation in the first place. Cleaning data is therefore an important step in analyzing it, and in many contexts, it may actually take the largest amount of time.

## Agenda

### Getting Started (5 mins)

**Survey Reminder - Last one!**

**Discuss: Why we need to clean data**

### Activity (40 mins)

**Clean Your Data**

### Wrap Up (5 mins)

**Reflection: Is data analysis objective?**

### Assessment

**Assessment Possibilities**

[View on Code Studio](#)

## Objectives

**Students will be able to:**

- Filter and sort a dataset using a spreadsheet tool.
- Identify and correct invalid values in a dataset with the aid of computational tools
- Justify the need to clean data prior to analyzing it with computational tools.

## Preparation

- Prepare data collected from survey to share with students. Ensure that a "Teacher only" master copy is kept safely somewhere.
- Student partners will carry through the next lesson and Practice PT. You may wish to select these pairs beforehand.
- Review the **Data Tools Resources** for this lesson (including Excel support)

# Teaching Guide

## Getting Started (5 mins)

### Survey Reminder - Last one!

**Survey Reminder:** Do one more entry in the class data tracker. You'll be using this data today. Give students a few minutes to fill out the class tracker survey that you started in **Lesson 7 - Introduction to Data**.

#### Teaching Tip

If you need to prepare the data ahead of time, you might not be able to squeeze in one last entry.

### Discuss: Why we need to clean data

#### Remarks

We have been collecting data about ourselves for several days. Now it's time to look at that data and see if we can find any interesting patterns or trends within it.

#### Discussion

Introduce the activity of the day and motivate the need to clean data before using it for analysis.

#### Prompt:

**"Before we get started, what challenges do you think we'll encounter as we begin to peek into the data we've been collecting?"**

#### Discuss:

Ask students to share their ideas with small groups or as a class.

- While there are presumably many challenges that will be mentioned, likely some of the comments will be related to the state of the data that was collected - in other words, how "clean" it is for analysis.

#### Transitional Remarks

There are many challenges associated with analyzing data. Today we're going to look at one that a lot of people don't often think about. When we collect data, it's usually "dirty," which means that, for one reason or another, it's not ready for analysis. We're going to investigate what this looks like and learn to use some tools to help us look at and "clean" the data.

## Activity (40 mins)

### Clean Your Data

#### Place Students in Pairs:

Students will clean and categorize their data in pairs. They will be using this data that they cleaned later in the unit for the practice PT.

#### Sharing the Data:

Pairs are going to need their own copy of the data collected from the survey. You should make your own master copy that will not be changed. To share the data with students, you can:

- Send a copy by email.
- Post a link to a Google Spreadsheet (make sure it's "View Only").
- **Note:** Instructions in Code Studio explain to students how they can "Make a copy" of a Google Sheet for themselves. If you are using a different spreadsheet tool, you should still share a copy.

#### Transition to Code Studio: Cleaning Data

Students will be guided through a series of activities that walk them through filtering, sorting, cleaning, and categorizing data.

The activity should be done in three parts.

### 1. Familiarizing Yourself with the Data:

Students learn how to sort and filter in a spreadsheet tool. There is no need yet to actually change any of the values. They simply should learn how these tools work in the spreadsheet tool you are using. Students can move on when they know how to filter and sort data.

### 2. Cleaning the Data:

Ignore “freeform text” responses for now -- for example, the “What did you do to relax?” column -- and focus attention on values that should be numeric or single words. Students will use sorting and filtering to find invalid values and will either fix or delete them. Students can move on when they have cleaned all “non-freeform” columns.

### 3. Categorizing Data:

Now focus attention on “freeform text” columns. Students will need to manually create new columns that categorize the inputs. This is a necessary step in order to perform computation with the data but it won’t feel very “algorithmic.” They will need to make choices, which is fine and will be addressed in the wrap up. Students can move on when they have cleaned all “freeform” columns by creating new columns of categories.

#### 💡 Teaching Tip

- You may wish to work through these set of activities as a class.
- When using Google Sheets or other online spreadsheet tools, it is possible for two students to clean the same dataset at the same time.
- Students should consult with their partners as they make their categorizations. Remind them that the goal is to have something they could analyze or chart later.

## Wrap Up (5 mins)

### Reflection: Is data analysis objective?

**Prompt: (Also found on Code Studio)**

**“In order to analyze data with a computer, we need to clean the data first. Based on your experience today, would you say that data analysis is a perfectly objective process? Why or why not?”**

#### 💬 Discussion

Students should reflect on the often subjective nature of cleaning data. Even as data is being cleaned to be used by computers, there will often be a “human element” to how it is cleaned.

#### 💬 Discuss:

Students should share their ideas in small groups before discussing as a class. The key ideas to touch on are:

- Data cleaning usually requires a human to make decisions about the data.
- There often will not be one “right” way to clean the data and different people will do it differently.
- Any categorizing in particular is quite subjective.

**NOTE: Make sure to save the cleaned up data:**

Pairs should save their data somewhere they can both access it. They will be using it in the following lesson.

## Assessment

### Assessment Possibilities

- Score or review a written response to the reflection prompt from the wrap up “Is data analysis objective?” (also found in code studio)

- Make a simple rubric (a checklist basically) for the steps of the activity that students were supposed to go through:
  - Used sorting in a spreadsheet
  - Used filtering to help identify outliers for cleaning
  - **Added a column** to categorize some form of free form text.

**Multiple Choice** (also on code studio)

**Which of the following is the most accurate statement about cleaning and filtering data?**

- Using computing tools to filter and clean raw data makes it impossible to analyze or draw accurate conclusions
- Filtering and cleaning data is a fully automated process that should not require human input or intervention
- Filtering and cleaning data is a human process that does not require the use of computers
- Filtering and cleaning data is necessary to ensure that data is in a form that is better for computers to process

## Standards Alignment

### Computer Science Principles

- ▶ **1.2** - Computing enables people to use creative development processes to create computational artifacts for creative expression or to solve a problem.
- ▶ **3.1** - People use computer programs to process information to gain insight and knowledge.
- ▶ **3.2** - Computing facilitates exploration and the discovery of connections in information.
- ▶ **7.1** - Computing enhances communication, interaction, and cognition.

### CSTA K-12 Computer Science Standards (2017)

- ▶ **DA** - Data & Analysis
- ▶ **IC** - Impacts of Computing



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: Creating Summary Tables

External Tools | Artifact Creation | Analyzing

## Overview

In this lesson students learn how to create their own summary tables from raw data. A summary table typically represents one or more aggregations (groupings of items) and computations that are performed on the raw dataset. In most spreadsheet programs, a summary table is called a pivot table. In the lesson, students learn how to make pivot tables in Google Sheets using a provided dataset. Then students turn to the data they've collected as a class and, with their partner, use pivot tables to investigate it further.

## Purpose

Making a summary (pivot) table is often considered an advanced technique. Once you get used to it, however, it's an extremely powerful computational tool that is available in most spreadsheet software. The purpose here is to acquaint students with using such a tool and to expose this power. Also creating summary tables is a direct tie to the CSP Framework essential knowledge statement: **3.1.3C Summaries of data analyzed computationally can be effective in communicating insight and knowledge gained from digitally represented information.**

The other purpose here is that creating a summary table is a good example of making a computational artifact for the **Explore Performance Task**. For that performance task students might find some raw data while doing research and might create a new artifact that is a summary table of the data that reveals some interesting aspect of it. Using a tool like a spreadsheet to make summary tables lets you explore data in deep ways, quickly and easily.

Being able to manipulate data is an important skill for computer scientists. Being able to create summary tables from larger datasets represents a form of computational thinking. To make a good summary table, one must have a good sense of the data, be able to hypothesize about what might be interesting to look at, and then have the skills to use a computational tool to create it. While seemingly mundane, a spreadsheet is an extremely powerful tool for working with data. Understanding the features of a spreadsheet tool, and what kinds of computations it can perform, can save you a lot of time and energy from either doing such things "by hand" or writing your own program to do it.

## Agenda

Getting Started

[View on Code Studio](#)

## Objectives

**Students will be able to:**

- Create a pivot table with at least one aggregation and one calculation when given a set of data.
- Describe the benefits a summary table has over a raw dataset.
- Collaboratively investigate a dataset by creating summary tables.
- Explain the meaning of a summary table they created.

## Preparation

- Review **Data Tools Resources** (including Excel support)
- Familiarize yourself with the tutorials about making pivot tables in Code Studio.
- Ensure students have access to the dataset they cleaned in the previous lesson.

## Vocabulary

- **Aggregation** - a computation in which rows from a data set are grouped together and used to compute a single value of more significant meaning or measurement. Common aggregations include: Average, Count, Sum, Max, Median, etc.
- **Pivot Table** - in most spreadsheet software it is the name of the tool used to create summary tables.
- **Summary Table** - a table that shows the results of aggregations performed on data from a larger data set, hence a "summary" of larger data. Spreadsheet software typically calls them "pivot tables".

**The need to create summary tables of raw data**

**Activity (90 mins)**

**Transition to Code Studio**

**Making Pivot Tables Part 1 - The Basics**

**Making Pivot Tables Part 2 - Manipulation and  
Visualization**

**Free play - make a summary table of the class tracker  
data**

**Wrap Up**

**Share and compare**

**Assessment**

**Assessment Possibilities**

# Teaching Guide

## Getting Started

### The need to create summary tables of raw data

#### 📍 Remarks

In the previous lesson we cleaned up the data we've been collecting. Now the question is: what can we do with it? Look at this table. It was created from the over 65,000 rows of data in the movie rating dataset we saw a few lessons ago....

#### 💡 Teaching Tip

As an alternative, you could show the little summary table above and ask the students:

- How long do you think it would take you to calculate the values in this table from the raw dataset of ~65,000 rows?

Based on what students know so far, they should guess relatively large amounts of time (dozens of minutes, or even hours). You can then reveal that today we'll learn how to make a table like this in roughly 10 seconds.

	Women		Men	
	Number	Avg. Rating	Number	Avg Rating
All Movies	16,716	3.54	48,819	3.53
Star Wars	102	4.23	284	4.37
Abyss, The	20	4.00	82	3.55

**This is an example of a summary table.** A lot of work and computation went into this. Notice that this is actually new data that was computed from the raw data. This is way beyond filtering and sorting. Computing this by hand for ~65,000 ratings (or writing formulas in a spreadsheet) would be pretty painstaking.

**But we can use computing tools to create summary tables like this for us in a flash** . Most data manipulation tools, like spreadsheets, allow you to quickly group, categorize, count, and average things. Making a summary table is a computational technique for exploring the data; let's try it.

## Activity (90 mins)

### 🖥️ Transition to Code Studio

**Put students back together with their data cleaning partner.**

- Students should go through the tutorial individually on their own computer, but should be seated next to their partner.

There are 3 levels in Code Studio that students go through.

Look at the levels in Code Studio for full details.

Here is a synopsis of what the students are being asked to do:

### Making Pivot Tables Part 1 - The Basics

The first tutorial walks students through the entire process of making simple pivot table using a provided data set.

Here are the steps they go through.

### Getting Started - Copy the Data

- A data set of movie ratings is provided.

### Your First Summary Table

- Select all the data and make a new Pivot Table

### Add Rows and Values to Your Table

- Organize the summary by listing each movie on its own row and show the average rating for it in each column.

### Summarize by: COUNT

- Change the value from the Average rating the COUNT of the number of ratings for each movie.

### Add Another Field to Values

- Add another column so that the table shows both the average rating and the count side by side.

### Teaching Tip

As you circulate the room, keep in mind the **key ideas** we want students to have:

- Summary tables (pivot tables) provide a way to visualize data
- summary tables allow you to see things in the data you might otherwise not see.
- Summary tables allow you to manipulate and create new data.
- A summary table helps you look at your data in new ways.
- A summary table can be a first step toward a good visualization.

## Making Pivot Tables Part 2 - Manipulation and Visualization

Students learn about a few more advanced features of making pivot tables and build up toward making a chart (visualization) based on a pivot table that they made, still using the movie rating data.

Here are the steps they go through:

### Adding Columns

- Add more columns (values) to the table to show more stuff for each movie

### Filtering Pivot Tables

- You can filter for values in a pivot table, just a like a spreadsheet - only show values that meet some criteria

### The Next Step - Manipulating the Pivot Table

- Copy the pivot table to a new spreadsheet in order to manipulate the values further -- once you have the basic table you want, manipulating it further in "pivot table mode" can be cumbersome since the computer needs to recompute the data every time you do anything.

### Moving on: Visualizing Summary Tables

- Make a chart of the pivot table you just made. See the examples in the tutorial on code studio.

## Free play - make a summary table of the class tracker data

The entire lesson builds toward students being able to make a pivot table of their own data - the data they cleaned previously.

### PLEASE NOTE: FREE PLAY is OPTIONAL

This free play should be considered optional or a bonus for this lesson.

If students have finished the tutorial they are ready to start the performance task in the next lesson.

## Wrap Up

Question: Did anyone find the potential makings of a data story today?

## Share and compare

- Have pairs of students share the pivot tables that they made with another pair, or with you, or with the whole class.
- This might be an opportunity for them to do peer-review of other groups' tables (see assessment below).
- Students should be able to describe what their table is showing, and preferably point out some insight they had.

**Recall the key ideas of summary tables:**

- **Summary tables (pivot tables) provide a way to visualize data.**
  - Yes, it's still a table, but by aggregating and summarizing information from a large dataset, summary tables allow you to see things in the data you might otherwise not see.
- **Summary tables allow you to manipulate and create new data.**
  - Even for our simple movies example here, the raw data didn't contain the average rating for every movie, or count how many ratings there were. We had to compute it, and the pivot table let us do that quickly and easily.
- **A summary table helps you look at your data in new ways.**
  - Think: how could data be grouped? What could be calculated? Once you know how to make a summary table you can begin to look at raw data and ask questions that you know might be possible to answer.
- **A summary table can be a first step toward a good visualization**
  - Often it's difficult to make a meaningful chart or graphic out of raw data. You often want to summarize it first, then chart it!

**Foreshadow:**

- In the next lesson, you and your partner will dig deeper into the data to find your own data story and make visualizations to tell it!

## Assessment

### Assessment Possibilities

**Note:** Formally assessing the pivot tables that come from this lesson should be considered optional. These partners will be making more pivot tables and charts from them for the Practice Performance Task in the next lesson.

**Multiple Choice:** (Also found in code studio)

Which of the following statements are true about pivot tables?

Select two answers.

- Pivot tables are used to quickly remove errors and inconsistencies from a dataset.
- Pivot tables are used to quickly perform aggregate computations and groupings on a set of raw data
- Pivot tables are used because they automatically detect and highlight potential trends or patterns in the underlying raw data
- Pivot tables are used to generate a summarized view of a large dataset which is helpful for gaining insight

## Standards Alignment

### Computer Science Principles

- ▶ **1.1** - Creative development can be an essential process for creating computational artifacts.
- ▶ **1.2** - Computing enables people to use creative development processes to create computational artifacts for creative expression or to solve a problem.
- ▶ **3.1** - People use computer programs to process information to gain insight and knowledge.
- ▶ **3.2** - Computing facilitates exploration and the discovery of connections in information.

### CSTA K-12 Computer Science Standards (2017)

- ▶ **DA** - Data & Analysis
  - ▶ **IC** - Impacts of Computing
-



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 7: Practice PT - Tell a Data Story

Practice PT | External Tools | Artifact Creation | Project

## Overview

For this Practice PT students will analyze the data that they have been collecting as a class in order to demonstrate their ability to discover, visualize, and present a trend or pattern they find in the data. Leading up to this lesson, students will have been working in pairs to clean and summarize their data. Students should complete this project individually but can get feedback on their ideas from their data-cleaning partner.

**Note:** This is NOT the official AP® Performance Task that will be submitted as part of the Advanced Placement exam; it is a practice activity intended to prepare students for some portions of their individual performance at a later time.

## Purpose

Students in this lesson will be telling their own story with a set of data about themselves. The hope is that using personal data will both motivate the exploration of the dataset and provide students with intuitions about the kinds of patterns or trends to explore. This Practice PT reflects many of the practices students will need to use on the actual AP® Performance Tasks, in particular the Explore PT. On that PT, students will need to create an artifact with a computational tool and explain both how it was created and what it is showing.

While students will not be required to create a chart or even necessarily visualize data for the PT, creating a data visualization would make for a strong computational artifact. This activity is designed to provide practice with one way to complete that aspect of the PT. Additionally students should leave this Practice PT familiar with many of the Learning Objectives related to the challenges of manipulating and analyzing data, as they will have now gone through the lifecycle of collecting, cleaning, analyzing, and visualizing data themselves.

**AP® is a trademark registered and/or owned by the College Board, which was not involved in the production of, and does not endorse, this curriculum.**

## Agenda

### Getting Started

Introduce the aims and goals of the Practice PT.

Activity (up to 3 days)

[View on Code Studio](#)

## Objectives

**Students will be able to:**

- Create summaries of a dataset using a pivot table.
- Manipulate and clean data in order to prepare it for analysis.
- Explain the process used to create a visualization.
- Design a visualization that clearly presents a trend, pattern, or relationship within a dataset.
- Create visualizations of a dataset in order to discover trends and patterns.
- Draw conclusions from the contents of a data visualization.

## Links

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

**For the Students**

- [Practice PT - Tell a Data Story](#) - Rubric

[Make a Copy](#)

**Create Individual Copies of the Data**

**Identify a Story**

**Visualize Your Story**

**Complete Written Responses**

**Wrap Up**

**Submit Practice PT**

**Sharing Work**

**Assessment**

**Rubric**

# Teaching Guide

## Getting Started

### Introduce the aims and goals of the Practice PT.

#### **Remarks**

Throughout this unit we have been collecting data about ourselves in the hope that we'll be able to find some interesting trends and patterns in that data. Today we're going to finally be able to take a close look at the data we've collected. Your job will be to use your new skills at cleaning, summarizing, and visualizing data to "tell a story" using the data we collected. We hope that, with so many different perspectives in the class, a lot of interesting stories on the same dataset will emerge.

## Activity (up to 3 days)

A proposed schedule of the steps of this project is included in the Teaching Tip as well as more thorough explanations of how to conduct the various stages.

#### **Practice PT: Tell a Data Story**

##### **Distribute:**

Distribute copies of **Practice PT - Tell a Data Story - Rubric** rubric and overview to students and review as a class. You may wish to read through the guidelines of the project together.

Alternatively, consider distributing the overview earlier in the unit to provide students an opportunity to preview and prepare.

Below are the steps of the PT that are laid out in the activity guide:

### Create Individual Copies of the Data

Students will have been cleaning and summarizing a shared copy of their data thus far. Now they should make separate copies of their data to complete this project. In Google Sheets, one student will need to go to "File" → "Make a copy" In Excel, one student can email the other a copy of their cleaned data

### Identify a Story

Students should already have some experience summarizing their data with a pivot table and visualizing it with charts. They should continue to iteratively use these tools to identify an interesting trend, pattern, or relationship within their data. Some good things to remind students:

- There's no need to tell a complex story. Simple relationships are still valuable to understand.
- The absence of a trend or pattern can still be interesting. If the amount of sleep you get doesn't have a clear impact on mood, that's interesting to know.

### Visualize Your Story

#### **Pacing Suggestion**

Below is a suggested timeline for completing the PT entirely in class. It's possible that this could be done in a single day to get most of the work done, with written responses being done for homework.

##### **Day 1**

- Students create individual copies of their data.
- Students summarize and visualize data, looking for an interesting story to tell.

##### **Day 2**

- Students identify a story in their data.
- Students design a visualization showing their data story.

##### **Day 3**

- Students complete written responses and submit their Practice PT.

Students should once again refer to the **Data Visualization 101** guide for tips on how to make clear visualizations. Their chart will have accompanying explanations, but it should be able to “stand on its own” to communicate the story students have found. Some good things to remind students of:

- A fancy chart may actually be worse than a simple and clear one.
- Creating multiple charts is totally appropriate if they will better communicate the story.
- Experiment with different chart types. The chart type used to discover the story may not actually be the best one for visualizing the story.

## Complete Written Responses

In the Practice PT, students will find responses modeled after those that will appear in the actual AP® Performance Tasks.

## Wrap Up

### Submit Practice PT

Students will need to submit their visualization and written responses. Direct students to check the rubric prior to submission to ensure they have all the necessary components.

### Sharing Work

As an optional addition to this project, have students share their findings. The visualizations can be placed around the room for a gallery walk, added to a single shared folder, or presented to the class. This is a good opportunity to see how different groups cleaned and interpreted the same dataset.

## Assessment

### Rubric

Use the provided **Practice PT - Tell a Data Story - Rubric** rubric, or one of your own creation, to assess students' submissions.

## Standards Alignment

### Computer Science Principles

- ▶ **1.2** - Computing enables people to use creative development processes to create computational artifacts for creative expression or to solve a problem.
- ▶ **3.1** - People use computer programs to process information to gain insight and knowledge.
- ▶ **7.3** - Computing has a global affect -- both beneficial and harmful -- on people and society.

### CSTA K-12 Computer Science Standards (2017)

- ▶ **DA** - Data & Analysis
- ▶ **IC** - Impacts of Computing



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**.