

# Pop Culture Data Science

## Lesson Overview:

This Hour of Code activity enables students to contribute anonymous opinion data on various student-relevant trends then visualize and analyze data by location and age to determine what is trending in different parts of the world.

### Lesson Summary

DURATION: 50-60 mins

#### Getting Started: (8-10 mins)

- Introduce the activity and crowdsourcing
- Pop Culture voting activity

#### Data Visualization Activity: (30 mins)

- Introduce Data Visualization in iSENSE
- Practice working with logical operators
- Facilitate and support students to complete Data Quests

#### Wrap-up Discussion: (10-15 mins)

- Discussion and Debrief

#### Extended Learning: (additional 60 minutes - optional)

- Optional extension activity

## Audience:

This lesson plan is intended for use with students in grade 6 and up (11 years old and up). It can be used to introduce Data Science and/or reinforce data analysis in a math or science course.

## Learning Objectives:

This learning objectives for this activity are to

- 1) Learn the concept of crowdsourcing and digital participation
- 2) Learn to use computational tools and techniques to collect and transform data to help others better understand real-world phenomena.
- 3) Learn to use visualization and analysis tools to find trends in data
- 4) Discuss the issue of bias in making predictions based on data sets and accessibility in the design of computational technologies that gather information.
- 5) Discuss the tradeoffs between allowing information to be public and keeping information private.

## Materials, Resources, and Preparation

- [Register your Hour of Code](#) event and promote it with some [fun posters](#).
- Secure a few Android devices (phones or tablets) for use in the beginning of the lesson.
- Be sure to test the [Pop Culture app](#) on an Android device and [iSENSE project](#) on a computer with display screen first before asking your students to use them. Check your technology and decide if you need to troubleshoot anything in advance.
- Run through the activity on using iSENSE to select and analyze data.
- Set up a projector to show students how to use the iSENSE project.
- [Print certificates](#) to hand out at the end.
- Recruit an older student or adult to serve as a facilitator who circulates and supports student problem solving.

## Getting Started (8-10 minutes)

### Directions on introducing the activity:

Kick off this Hour of Code activity by inspiring students to get involved in crowdsourced projects and in conducting data analysis to learn about other like them.

#### Introduce Crowdsourcing:

What is crowdsourcing? Does anyone have any experience with it? If so, describe the experience and say why you participated.

Cite the definition (from wikipedia):

Crowdsourcing is the practice of obtaining information or input into a project by enlisting the services of a large number of people typically via the Internet. *Example:* Wikipedia. Wikipedia gave a crowd the ability to create an online encyclopedia on their own.

Why do I need to know about crowdsourcing:

In our vastly interconnected world, crowdsourcing has been used to share information and tasks such that more voices could contribute and the work could be divided among many.

Crowdsourcing has been used in science to generate data sets and to enlist many people and computers in search for solutions. *Example:* FoldIt is a revolutionary crowdsourcing computer game enabling you to contribute to important scientific research.

Why teach crowdsourcing as part of a CS curriculum?

Crowdsourcing is explicitly addressed as a goal in the CSTA standards. 2-IC-22 6-8th

“Collaborate with many contributors through strategies such as crowdsourcing or surveys when creating a computational artifact.”)

Introduce the Pop Culture activity:

In this activity, you may contribute your anonymous opinions through the Pop Culture app.

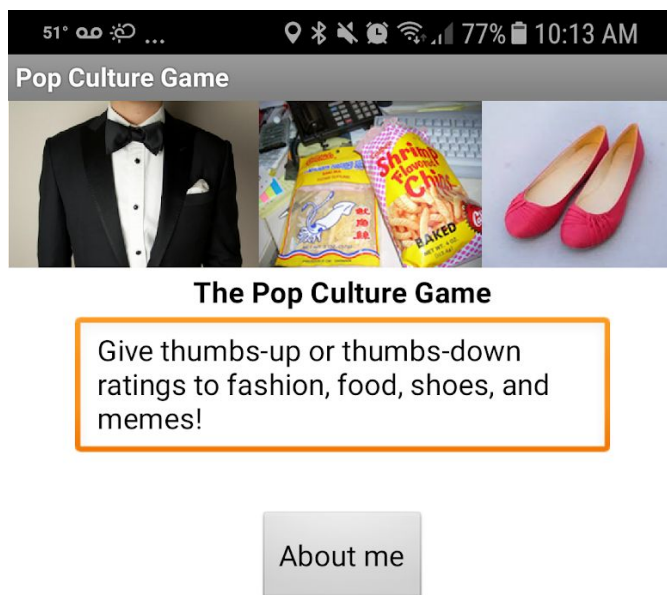
Open the Pop Culture App Inventor app by clicking on this link from an Android phone

[https://play.google.com/store/apps/details?id=appinventor.ai\\_fgmartin13.Pop\\_Culture](https://play.google.com/store/apps/details?id=appinventor.ai_fgmartin13.Pop_Culture)

(direct APK download is here

<https://drive.google.com/open?id=0B41atXIV0KHKbUNldWlwa2NMcGM>)

The opening screen describes the Pop Culture game.



Click the “About me” button then enter your information:

- Gender (“him”, “her”, or “other” - allowing for a gender neutral response)
- Age (note that selecting a range will default to the average age in that range)
- Location (If the phone’s GPS data is accessible, a location is set with a bit of randomness to obscure the actual location; or if GPS is not available, you pick your State and what gets logged is a location close to the center of that State.)

## About\_Me



Tell us a bit about yourself, and then start rating memes, food, and fashion.

### 1. My gender



Her



Him



Other

2. My age 15

3. My location set from GPS

4. Let's play!

Click on "Let's Play!" to continue.

Game play:

In this segment, you view items one at a time and give a thumbs up or thumbs down rating to show if you like or dislike the item. After each vote, you will be advanced to the next item.

Thumbs\_Up\_or\_Thumbs\_Down?



***Fresh Fruit***

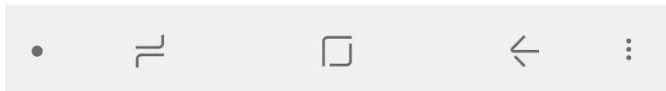
11 of 54



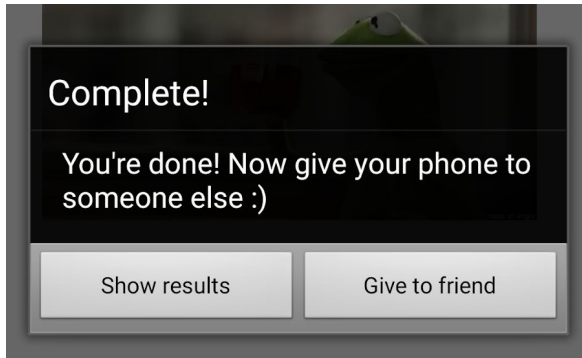
0



1



Upon completion, you will be asked to pass along the app to another student so they can vote or choose to show the results of your votes.



Give students 5 more minutes to finish voting. Try to minimize the cross talk by stating that we will have a time to compare our opinions, as well as those of many others, later in the activity.

### Introducing Data Visualization in iSENSE

“Show results” opens up the iSENSE extension. We highly recommend projecting iSENSE from a computer screen. View the iSENSE data at [http://isenseproject.org/projects/3177/data\\_sets](http://isenseproject.org/projects/3177/data_sets) .

The data you entered, as well as data from other players, are viewable in iSENSE. Along the left hand side you will see a set of tools.



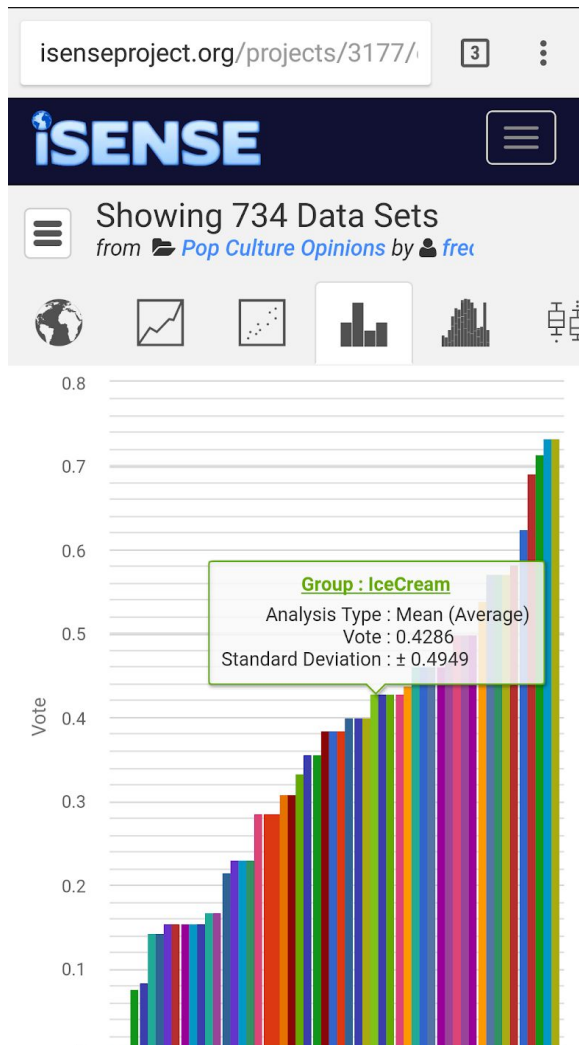
**Tap to the right of the tools panel to close the panel, OR if on the desktop, click on the**



triple bar (also called the hamburger) to the right to close the tools panel.

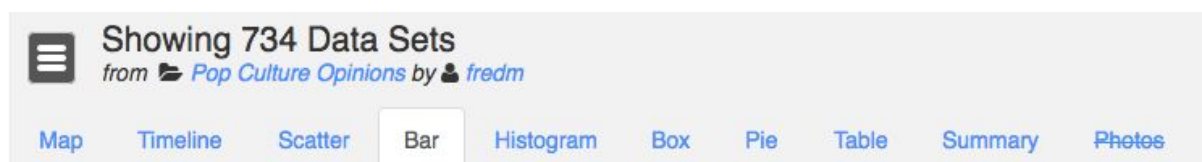
With the tool panels minimized you can see a visualization of the data. The visualization panel shows different views on the data set.

The Bar Graph is currently selected and the data look like this...



When you tap on each bar, you can see its name and average rating this item received. A perfect score is 1.0 (all thumbs-up) and 0 is all thumbs-down.

Clicking on Map from the visualization tool bar, you can see a map of where people contributed data.



## **Working with the Visualization Tools**

In the panel across the top of the page, you can click on the tabs to view the different types of graphs. If we click on the “Summary” tab, we will see an overview of some of the important data related to your project.

### **Working with Maps**

#### **Introduction**

In the “Maps” tab you can view the latitudes and longitudes where people contributed data to this project. You can zoom in and out using the plus and minus buttons at the bottom right corner. You can drag the map to peek around.

#### **Customizations**

The map has markers all over it for each of your data points. Click on one to view all its information. If there are many at the same location, they will spread out for you when you click so that you can see them all. If you’re too zoomed out, the data points might be lumped together into a single number representing the number of data points at that location.

### **Annotation**

#### **Commenting**

To add a comment to a point, all you have to do is select the point, and click the plus button. You’ll get a box where you can enter a message. Just press the checkmark to finish, and it’ll show up. If you don’t have a point selected, you can still add a comment. Again, click , and enter a message. The comment comes up in the corner, but it can be dragged all around. Zoom out, and it stays right in place for you.

#### **Editing**

You can edit comments after you make them. Select a comment by clicking on it, then click the button. You have some formatting options, such as bold, italic, and using large font. You can also create a link that makes the text a clickable button. Press the checkmark to save, or the trashcan to delete the comment. For points, just select the point to get the edit option.

#### **Saving**

Your comments won’t be around anymore when you finish your session. To keep your comments, save a visualization, just like we always do on iSENSE when we want to save a layout.

### **Working with Bar Graphs**

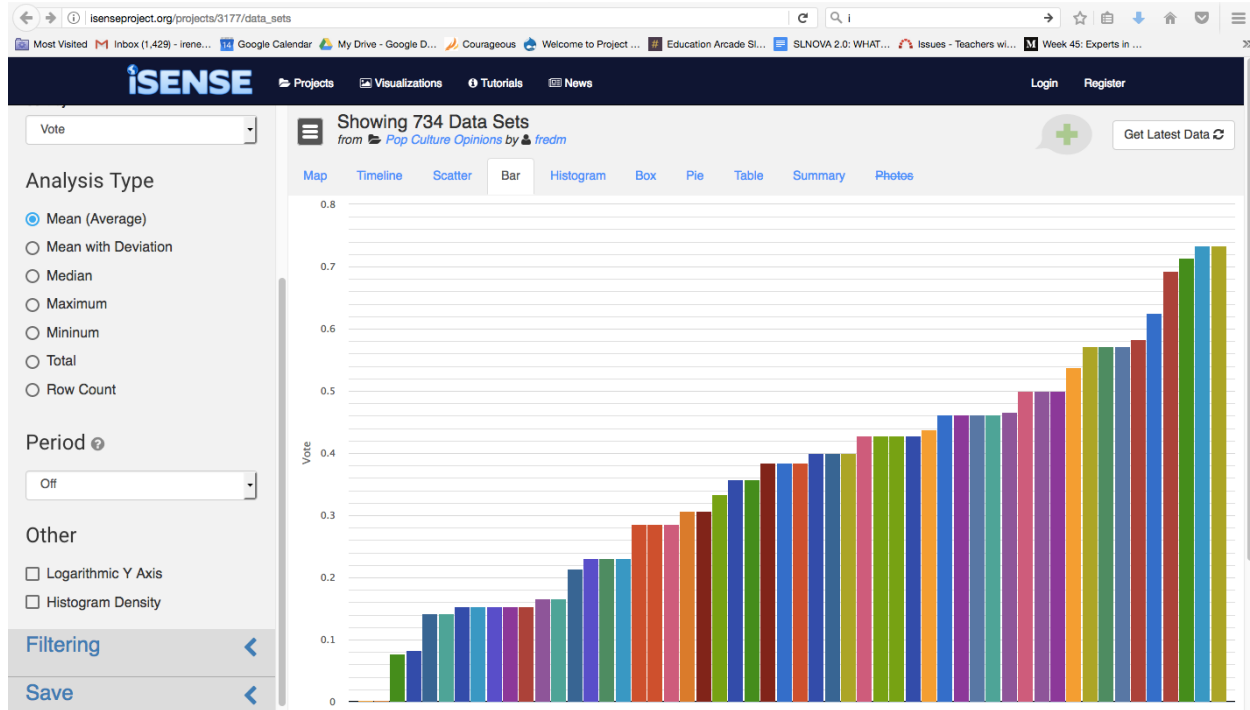
#### **Introduction**

Let’s take a look at the “Bar” tab.

#### **Customizations**

If you have more than one group, they will all show up on the graph. Click on “Groups” to choose which ones to hide or to change the color of a bar. Under “Y-Axis” we can control which data the bars represent. Here, for instance, we have the average votes on the Y-Axis, so we can see that the baggy jeans had a low average number of votes while hoodies had a high average number of votes.





## Tools

Let's take a look under "Tools". Right now, the bars are showing us the average votes. This tells us that among the voters fresh fruit and nuts were popular.

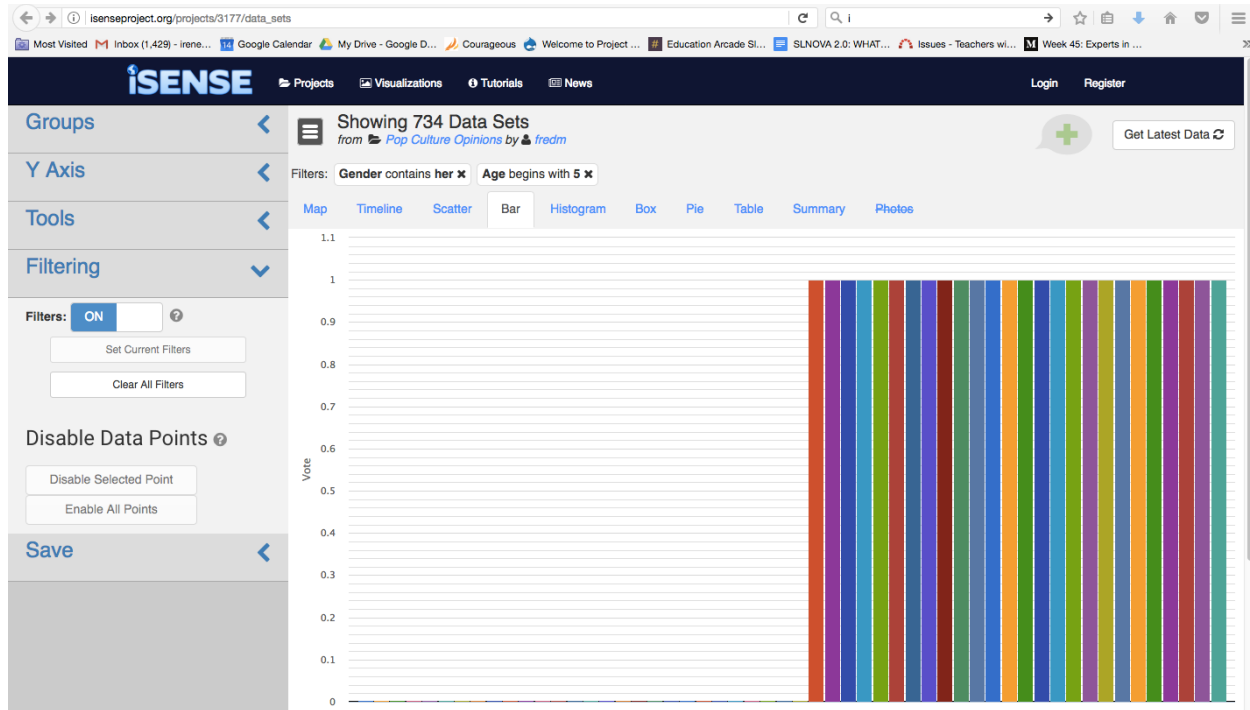
## Demo

Now let's get into the demo. Say I'm going over to my aunt's house and was told to bring a snack but I have no idea what to bring. I could query this database to find out what females over 50 liked to eat as snack foods. How would I do this? See the next section on Working with Filters...

## Working with Filters

### Adding a Filter

The first step to using filters is to turn on the feature in the visualization page. (Turn on the feature by opening up the "Filtering" tab on the left hand side tools panel. Then click on the ON part of the toggle.) Filters let you choose which data you want to see in your visualizations and which you don't. The first important thing about filters is that when you turn them on, they filter *every* type of chart, not just one. So, from our previous example, we want to find out what snack foods people over 50 like. Click over to the table tab and find the age column. Next click on the "~" to select which data to keep. Since we want to focus on data with age around 50 in the age field, we could choose "starts with" and type "5" into the field. We can also choose "gender" to be "her". Click "set current filters" to narrow the input set according to your specifications.



From this analysis we predict that my aunt might like shrimp chips or sushi.

(Optional discussion: What is wrong with using this data for making predictions?)

## Removing a Filter

To remove a filter, click the “X” next to it. Or, click “Clear” to delete them all at once.

## Working with Tables

### Introduction

“Tables” are the simplest visualization. The table view is simply a big list of all your data in which each row represents all the numbers associated with a single data point. In our project, [http://isenseproject.org/projects/3177/data\\_sets](http://isenseproject.org/projects/3177/data_sets), the columns represent the data point ID, the data set name, the contributor, the timestamp, lat, long, gender, age, item name, and vote. Make sure you have “#Select all” checked in the Visible Fields tools on the left hand panel.

### Customizations

These default columns we just discussed can all be checked on and off under the “Visible Fields” section. You can also check your own fields on and off under this section, to control whether or not they show up in the table; use the “Groups” tab to restrict which groups show up in the table as well.

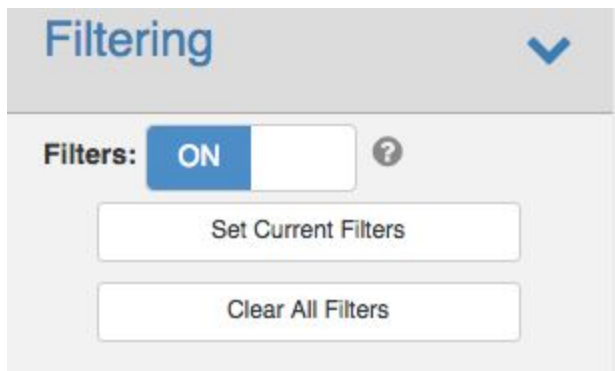
### Searching Through Data

Use the logical operators provided. For numerical values the logical operators provided are equal, not equal, less than or equal, greater than, and greater than or equal. For text fields: the logical operators provided are contains, does not contain, begins with, does not begin with, ends with, does not end with, is in, and is not in. These operators along with a value or string can be used to specify the characteristic you want to keep in the data set.

<b>==</b> equal
! not equal
< less
<= less or equal
> greater
>= greater or equal

<b>~</b> contains
!~ does not contain
^ begins with
!^ does not begin with
ends with
!@ does not end with
= is in
!= is not in

For example, in the previous example, age equal 50 and gender contains her was used to filter data to create a data set that only contains responses from females of age 50. Note; The subset of the data needs to be “set” by clicking on the “Set Current Filters” button on the Filtering tool on the Tools panel on the right hand side.



### Saving a Visualization (optional)

**To save a visualization** go to the “Save” tab, and click on the green “Save Visualization” button. (Note: you have to be logged in to submit a visualization to the library.) Your visualization, along with all the current filters and colors you added, will now be saved. To view your newly saved visualization, or any other user’s saved visualizations for that matter, go to the “Visualizations” page.

### **To download a snapshot of your visualization without saving.**

Also under the save tab, you can click “Download” to save a picture of your graph to your computer, or “Print” to print it.

### Practice with logical operators:

1) I want to find data submitted by 14 year old males, what logical expressions should I set up?

Answer: "age contains 14 and gender contains him"

Note that age is treated as a character string because of ranges.

2) I want to find out if people older than 19 like costuming more than dislike it, what logical expressions should I set up?

Answer: "age does not begin with 1 and thing contains costuming"

3) I want to find out if younger people like more things in general than older people, what logical expressions should I set up?

### Sample Data Quests for students:

1. After using the Android app to rate the 54 items, find your own replies on iSENSE. (Go to [isenseproject.org](http://isenseproject.org) and find the [Pop Culture Opinions project \[number 3177\]](#). Scroll down to the Data Sets area and find your data set by looking for the three initials and when you contributed it. Then click on your data set and you should see a bar graph with high bars for things you gave thumbs-up, and barely-there bars for things you gave thumbs-down.)
2. What are the most popular things? (How do you support your claim?) (Use Bar Chart; Group By > Thing; select Analysis Type > Mean; mouse over bars)
3. What are the least popular things? (How do you support your claim?) (Use Bar Chart; Group By > Thing; select Analysis Type > Mean; mouse over bars)
4. How many people have voted on these things? (Use Bar Chart; Group By > Thing; select Analysis Type > Row Count; mouse over bars)
5. What are the things other persons of your age and gender like most? (How do you support your claim?) (Go to Table; type your gender and age into the respective table columns; go to Filtering, turn it On and click Set Current Filters. Then, go back to Bar Chart, put Analysis Type on Mean, and mouse over bars.)
6. What are the things other persons of your age and gender like least? (How do you support your claim?)
7. Are French Fry likers primarily male or female or other? (How do you support your claim?)
8. Overall, are there more votes in favor of items, or against them? (histogram shows the answer)
9. Of him, hers, or others, which group is overall more positive about the items? (Group by Gender; use Bar Chart; set Y axis to Vote; set Analysis type to Mean)
10. What is the ratio of him, her, and others contributing votes? (Group by Gender; use Pie Chart; set Analysis Type to Row Count)

Coming soon .... In the next version, we will provide a way for the student to select a geographic regions and give it a label. Then the data can be filtered using those labels and comparison can be made on the popularity of items in one region compared to another region.

**Discussion:** (*recommendation:* have students discuss their responses with an elbow partner or in small groups, then ask each team to share out.)

Introduce the concept of prediction and how predictions from data impact our daily lives. A prediction is a statement about what will happen or might happen in the future. An example of how predictions impact our daily lives is that we will buy items we search for or have bought before online. Thus, ads that we are shown on the internet are selected for us based on our online viewing and past purchasing habits.

Q: Ask students to make predictions about what students like in other parts of the country?

Q: Do the data support the predictions? Why or why not?

(A: most likely, insufficient data from different regions of the country exists to make predictions)

Introduce the concept of bias and how bias can be intentional or accidental.

Q: Where in the process of developing the app could bias creep in?

(A: the data set itself may only contain data from a certain type of person or certain geographical location thus it may be tuned to some preferences and not others. The images chosen for the survey may be foreign to people in other cultures. etc.)

Were there pictures that people in other parts of the world might not understand?

Is the concept of liking and disliking foreign to other cultures and socio-economic groups?

Review the collection of images and discuss how each may be biased.

Where in the processes of collecting, organizing, and displaying visualizations might there be biases?

How could the Pop Culture app be redesigned or adapted to reduce bias?

Some groups may choose to use the data collection without adding their own, is this good or bad?

What are some of the tradeoffs between sharing your data and keeping it private?

### **Design an Investigation (optional extension activity):**

Design a question that you are interested in answering through data analysis. Develop your question, hypothesis, method and findings. Share your plans with others.

### **Standards addressed:**

#### **CSTA K-12 CS Standards (2017)**

2-CS-01 6th-8th Recommend improvements to the design of computing devices, based on an analysis of how users interact with the devices.

2-DA-07 6th-8th Represent data using multiple encoding schemes.

2-DA-08 6th-8th Collect data using computational tools and transform the data to make it more useful and reliable.

2-IC-20 6th-8th Compare tradeoffs associated with computing technologies that affect people's everyday activities and career options.

2-IC-21 6th-8th Discuss issues of bias and accessibility in the design of existing technologies.

2-IC-22 6th-8th Collaborate with many contributors through strategies such as crowdsourcing or surveys when creating a computational artifact.

2-IC-23 6th-8th Describe tradeoffs between allowing information to be public and keeping information private and secure.

### **Common Core Math Standards**

CCSS.Math.Content.6.SP.A.1

Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.

CCSS.Math.Content.6.SP.A.2

Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.

CCSS.Math.Content.6.SP.A.3

Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

CCSS.Math.Content.7.SP.A.1

Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population.