ACTIVITY 1
Data, Data, Data!

The ease with which data can be stored, cataloged, and searched in an ever more-connected society is an important point to understand. Data storage has both positive and negative implications, and being able to recognize these, as well as how your actions can contribute to the data being stored, are addressed in the following activity.

1. List two or three sites that you access on a regular basis.

   Site 1: ebay.com

   Site 2: youtube.com

2. In your group, find the privacy policy for your given site and identify two or three pieces of information that are collected for users of the site.
Chosen site: ebay
Information collected: purchase history

File Types

3. What does the word *delimited* mean? Why is this necessary when talking about data files?
A sequence of one or more numbers, this is how the data is stored and the coding software has to read it

4. Given a data file, how can you determine the type of data that might be contained in a specific column?

How the file is saved?

Posing Questions

5. On your own, identify two broad areas of interest that you have. Examples include health, sports, etc.
Interest 1: video games
Interest 2: cars

6. For each of the two areas of interest you identified, determine one question to which you might want to know the answer. These should be questions that are not easily answered with an online search. Two examples of questions that are easy to answer with an online search are who won the 1998 Superbowl (The Denver Broncos), and what is the height of the world's tallest building (at the time of publication, Burj Khalifa, 2717 ft/828 m).

Question 1: how many lines of code does cold war have

Question 2: best thing to change in your far for better performance

Discuss your two questions with your partner and refine your questions based on their feedback. Write your updated questions below.

Question 1: how many lines of code does cold war have

Question 2: best thing to change in your far for better performance

Finding Data Sets

These sites that provide free access to data sets. Spend a few minutes looking at the sites.

- http://www.kaggle.com (This site requires a free account, which can be an existing Google account, in order to access and search data sets.)
- http://www.data.gov
- https://toolbox.google.com/datasetsearch

7. Are there data sets that might apply to one of the two questions that you refined in question 6? Find two different data sets that might be used to answer one of your questions. List the site and any search criteria used to find the data sets. If not, consider revising one of your questions so that you can identify an applicable data set.

Site: Kaggle

Search criteria: Video game sales

8. How many records are in each of the data sets you identified? Describe one benefit of using a larger data set with more records.

Over 55 thousand different pieces of data, with a larger data set you get more accurate numbers that are closer the real value

Check Your Understanding

9. With your partner, discuss one way that user data captured by a site (whether knowingly or unknowingly) has contributed to an improvement in the service provided. Have there been any positive impacts of this data outside of the service or website?

It gives companies an idea of possible what users want to see and what users don't want to see in games, if most people buy games with a certain game mode then more companies will add that gamemode. Im sure there has been data collection is very beneficial to companies, services, and websites

10. Do you know how the data in the data set you identified was collected? If so, please describe. If not, describe one way that the data might have been collected.

11. In your opinion, are there situations where the benefit provided from the data collected is worth the risk to personal privacy? Why or why not?