

Big, Open, and Crowdsourced Data

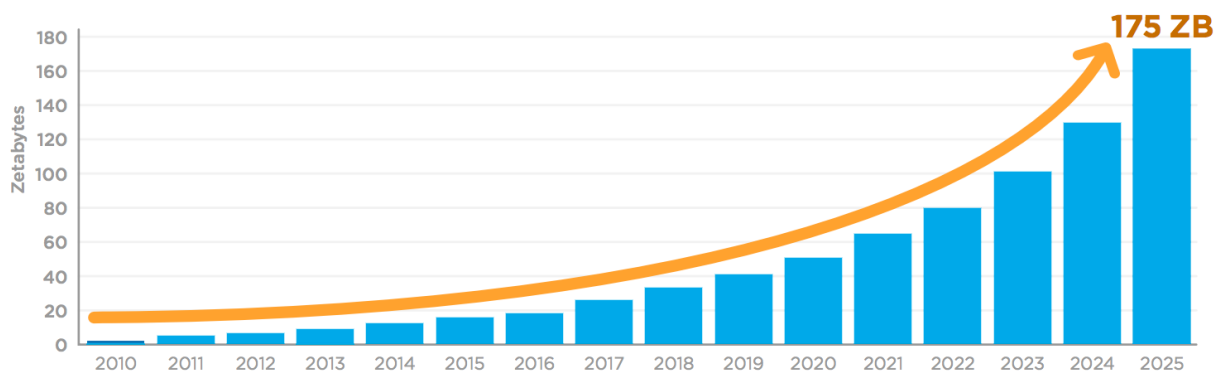
Your Tasks (Mark these off as you go)

- ☐ Explore Big Data
- ☐ Explore Crowdsourced Data
- ☐ Explore Open Data
- ☐ Define key vocabulary
- ☐ Receive credit for this lab guide

□ Explore Big Data

The digital world is constantly collecting more and more data. Whenever you use an online service, you're contributing to a data set of user behavior. Even by simply using electricity and water in your house, you're contributing to a data set of utilities usage.

With the increasing number of people and cities connected to the Internet, data sets are increasingly larger in size. One report estimates that the total size of digital data will be **175 zettabytes** in 2025.



How much data is 175 zettabytes, anyway? A single zettabyte is a trillion gigabytes. A modern smartphone stores about 32 gigabytes. To store 175 zettabytes, we would need 6 trillion smartphones (1000 smartphones for every living person!).

Whew, that's a lot! But how big are the individual data sets?
These stats can give us an idea...

- A single MRI scan results in **20,000 images**.¹¹
- Google processes **3.5 billion search queries** per day.²²
- Instagram users post **54,000 photos** each minute.³³
- An autonomous vehicle generates **11 terabytes of data** each day.⁴⁴
- Twitter users post **3,000 tweets** every second.⁵⁵

Big data sets are so large that our traditional ways of storing and processing them are no longer adequate, presenting challenges to computer scientists and data engineers. On the plus side, they're also so large that they offer new opportunities for analysis that were impossible on a small data set.

In this lesson, we'll explore where big data comes from and the exciting ways that we can use it.

Watch the videos below about Big Data. As you watch the videos keep notes on how the data analysis process can be applied to Big Data.



<https://www.youtube.com/watch?v=1XGo8K1boH4>



<https://www.youtube.com/watch?v=bMrDHtGHFR4>

Provide an example of big data from each of the videos.

Video 1:

Video 2:

Indicate how the data was filtered, cleaned, or processed to help solve a problem.

Video 1:

Video 2:

How is the data visualized or presented?

Video 1:

Video 2:

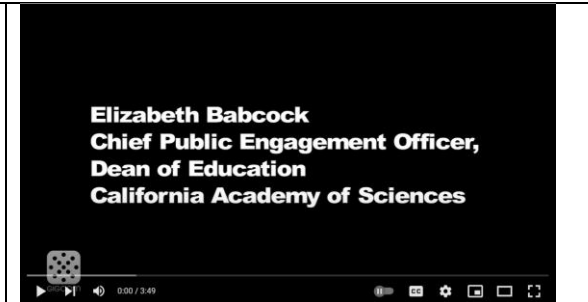
How is the new information gleaned being used to solve a problem or make decisions?

Video 1:

Video 2:

□ Explore Crowdsourced Data

Listen the NPR news snippet on How Pokemon Inspired A Citizen Science Project. Then watch the video on What is Citizen Science? As you listen to the news snippet and watch the video, keep notes on how the data analysis process can be applied to Citizen Science.

<p>ENVIRONMENT</p> <h3>< How Pokemon Inspired A Citizen Science Project To Monitor Tiny Streams</h3> <p>April 20, 2018 • 5:06 AM ET</p> <p>3-Minute Listen + PLAYLIST Download Share Menu</p>	
---	--

<https://www.npr.org/2018/04/20/597972310>

<https://youtu.be/81hhecl0p5k>

What is the source of data? Provide an example from the audio and video clip.

Audio Clip:

Video:

Provide examples of the data collected.

Audio Clip:

Video:

For each example above, indicate how the data collected could be visualized?

Audio Clip:

Video:

For each example above, how is the new information being used?

Audio Clip:

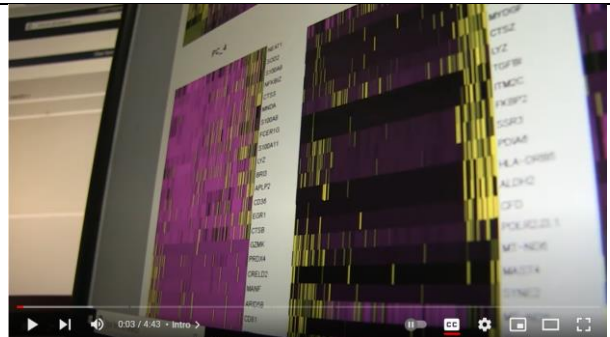
Video:

□ Explore Open Data

Watch the videos on Open Data below. As you watch, keep notes on how the data analysis process can be applied to Open Data.



<https://www.youtube.com/watch?v=qSD9ob8rGcs>



<https://www.youtube.com/watch?v=iOrPK7p2AwI>

Provide an example of open data from each of the videos.

Video 1:

Video 2:

Provide an example from each of the videos of how open data can be filtered to solve a problem.

Video 1:

Video 2:

For each example above, indicate how the data collected could be visualized?

Video 1:

Video 2:

For each video, provide an example of how Open Data can be used to solve a problem.

Video 1:

Video 2:

❑ Define key vocabulary

Scalability

Parallel Systems

Citizen Science

Crowdsource

Open Data

Open Access

❑ Receive Credit for this lab guide

Submit this portion of the lab to Pluska to receive credit for the lab guide.