

(<http://csmatters.org/pd-new>) C - 03

0bC - 0b11

# Manipulating Large Data Sets

## Unit Concept Lessons

**Revision Date:** Jan 05, 2018

**Duration:** 60 50-minute sessions



### Lesson Summary

**Summary:** Teachers will master Lesson 5-1, Manipulating Large Data Sets. They will learn concepts related to big data, including definite opportunities, limitations, and problems, and will review the lesson components. Finally, they will complete practice exam questions. Teachers will also reflect on the cognitive differences between declarative and procedural knowledge and why "knowing how" is an important learning objective for students.

#### Outcomes:

Teachers will:

- Explore both sessions of Lesson 5-1, Manipulating Large Data Sets.
- Complete the lesson's Session Two data analysis activity.
- Compare the CS Matters and the College Board perspectives on this topic.

#### Overview: (Total: 60 min)

1. Journal (10 min)
2. Big Data Questions (20 min)
3. Lesson 5-1 Review (20 min)
4. Practice Exam Questions (10 min)

### Learning Objectives

#### CSP Objectives

**Big Idea - Professional development provides ongoing opportunities for teachers to practice working with colleagues, including other teachers, principals, counselors, social workers, and others, and emphasizes that collaboration is a means and not an end in addressing issues related to school improvement and improved student learning.**

- ○ LO 3a - Professional development provides ongoing opportunities for teachers to practice working with colleagues, including other teachers, principals, counselors, social workers, and others, and emphasizes that collaboration is a means and not an end in addressing issues related to school improvement and improved student learning.

**Big Idea - Professional development matches learning experiences, including the intensity and duration, with individual teacher needs, current knowledge and skills, and learning goals.**

- ○ LO 9a - Professional development matches learning experiences, including the intensity and duration, with individual teacher needs, current knowledge and skills, and learning goals.

**Big Idea - Professional development combines a variety of learning experiences, including, but not limited to, individual study, demonstrations, observation, practice, feedback, and reflection as well as opportunities for collaboration and problem solving among colleagues.**

- ○ LO 9b - Professional development combines a variety of learning experiences, including, but not limited to, individual study, demonstrations, observation, practice, feedback, and reflection as well as opportunities for collaboration and problem solving among colleagues.

**Big Idea - Professional development relies on information technologies to provide more extensive and diverse content, and it also relies on communication technologies to expand access and participation and to create virtual professional learning communities.**

- ○ LO 9d - Professional development relies on information technologies to provide more extensive and diverse content, and it also relies on communication technologies to expand access and participation and to create virtual professional learning communities.

## Teacher Resources

Student computer usage for this lesson is: **required**

CIO3\_Manipulating Large Data Sets Folder

([https://drive.google.com/drive/u/2/folders/0ByBRH8unk6\\_5RDRfcTFEN2Vqa0k](https://drive.google.com/drive/u/2/folders/0ByBRH8unk6_5RDRfcTFEN2Vqa0k))

Manipulating Large Data Sets ppt

(<https://docs.google.com/presentation/d/18RxmTsSO4qp0kByuUAz0-G44Mh-IFZiF8Gm9YyrfW4/edit?usp=sharing>)

## Lesson Plan

TOTAL: 60 min

**Journal (10 min)**

## Questions

*Note: This section uses slides 2-3 of the powerpoint*

- What questions might a student ask about Big Data? Each teacher should write at least two such questions on their own.
- Work within table groups to find at least 5 such questions.

## Discussion

*Note: This section uses slides 4-5 of the powerpoint*

The instructor should discuss the rationale for accessing prior knowledge and the two knowledge types, declarative and procedural.

- Declarative knowledge can also be described as rote learning or “knowledge-telling,” which may include many facts and details that do not form an integrated whole. Students who have declarative knowledge are able to answer fairly simple reproduction tasks that do not require an ability to integrate or apply knowledge.
- Procedural knowledge, on the other hand, is characterized by an ability to integrate knowledge, understand relations between concepts, and, at the highest level, apply this knowledge to problem solving. It is often referred to as “knowing how” and is closely related to higher-order cognitive skills.

## Big Data Questions (20 min)

*Note: This section uses slides 6-14 of the powerpoint.*

Below are the main questions explored about big data in this section:

1. What is Big Data?
2. What opportunities does it offer?
3. What are its limitations?
4. What are its problems?

The instructor should:

- Discuss answers to each question.
- Discuss any questions left unanswered from the journal prompt.
- Present the College Board CSP and the CS Matters perspectives on Data Analysis.
- **Note:** read the speaker notes in each slide for more information on presenting the material

## Lesson 5-1 Review (20 min)

*Note: This section uses slides 15-19 of the powerpoint. In the third step, the instructor will need to walk the teachers through what the demonstration will look like.*

1. Present the Session 1 Lesson Outline.
  - The lesson has two sessions. Session One introduces what Big Data is through a reading and a video. (Show the video from 1:12 - 4:00.) The reading is available online but requires registration to access. You may want to download a pdf version of the document and make it available to your students for the jigsawed reading portion of the first session.
2. Present the Session 2 Lesson Outline

### 3. Demonstration

- Demonstrate steps 1 - 6, focusing particularly on steps 2 and 5.
- Extract data source into format supported by underlying tools.
  - Open bankfailure.csv in Google sheets and filter by state.
  - Attempt to open Consumer\_Complaints.csv in Google sheets. Explain the error and the need to move to alternate tools.
  - Open Consumer\_Complaints.csv in excel and apply filters.
- Perform analysis.
  - We have determined that the bank “Banco Popular de Puerto Rico” is on both lists. Now ask the students, “Why is this bank on both lists?”
  - **Note:** On the Failed Bank list, the Banco Popular de Puerto Rico is actually an acquiring institution. By looking more closely at the dates of the acquisition of the failed bank “Westernbank Puerto Rico,” one can formulate a hypothesis that perhaps the reason “Banco Popular de Puerto Rico” is on the complaint list is because they had recently taken over a failed bank. It could be possible that some of these complaints were related to this recent acquisition.

## Practice Exam Questions (10 min)

Complete the practice exam questions listed on slide 21.



(<http://www.umbc.edu/>)



(<http://www.umd.edu/>)



(<http://www.nsf.gov/>)

*Authored by:* CS Matters in Maryland

*Website:* [csmatters.org](http://csmatters.org) (<http://csmatters.org>)

*Email:* [csmattersinmaryland@gmail.com](mailto:csmattersinmaryland@gmail.com) (<mailto:csmattersinmaryland@gmail.com>)

This work is licensed under a  
Creative Commons Attribution-ShareAlike 3.0 United States License  
(<http://creativecommons.org/licenses/by-sa/3.0/us/>)

by University of Maryland, Baltimore County (<http://umbc.edu>) and University of Maryland, College Park  
(<http://umd.edu>).