API CAN CODE

COLLEGE OF EDUCATION

# Unit 1: Data in Learners' Lives

## Lesson 1.3: Using Data

*In this lesson, students will learn about the process of generating insight from data and will be introduced to the* Data-Information-Knowledge-Wisdom (DIKW) model. *As an example, they will discuss visualizations of data from the COVID-19 pandemic in Washington, D.C.*

**Duration:** 90 minutes

**Objective:** By the end of this lesson, students will understand the Data-Information-Knowledge-Wisdom (DIKW) model to make sense of the stages of a data science investigation.

**Lesson Walkthrough Video:** Unit 1 Lesson 3 - Teacher Walkthrough

### CSTA Standards in this Lesson

| Identifier | Concept | Subconcept | Standards |
|---|---|---|---|
| **HS-SYS-SE-32** | Systems & Security | Security | Classify the causes and impacts of security breaches and social engineering attacks for individuals, industries, communities, and governments. |

### CSTA Data Science Specialty Standards in this Lesson

| Identifier | Concept | Subconcept | Standards |
|---|---|---|---|
| **S1-DSC-MI-07** | Data Science | Interpreting Models, Problems, & Results | Interpret the results of data analyses to explain patterns, anomalies, and trends, and connect them back to the original problem or research question. |
| **S2-DSC-PP-14** | Data Science | Professional Practice | Evaluate protective measures in data collection, usage, and governance for privacy, security, and fairness. |

### Lesson activities

**Warmup (10 minutes)**

*(CSTA standards in this activity: 3A-IC-24, 3A-IC-29)*

- Read the following article about user data stolen from 23andMe: (https://news.yahoo.com/user-data-stolen-genetic-testing-203545299.html).
- Ask the students the following questions:
  - What stakeholders are involved? [23andMe is the main stakeholder]

This work was made possible through generous support from the National Science Foundation (Award # 2141655). 

- What issues of privacy are highlighted? [A group of hackers got into their system and stole user data, which is apparently now being sold on the dark web.]
- How could those involved have mitigated this risk? (or in other words, how could they have better protected user data privacy?) [Should 23andMe have stored user data beyond the time necessary to process it?]

**DIKW Lecture (20 minutes)**

*(CSTA standards in this activity: 3A-DA-11, 3A-DA-12)*

- The teacher will introduce the Data-Information-Knowledge-Wisdom model, first in general terms and then as applied to the investigation of the 1845 Cholera outbreak by John Snow, using the provided slides.
- The teacher will explain to the students that the Cholera epidemic is a widespread outbreak of a severe diarrheal disease caused by the bacterium Vibrio cholerae.

**Understanding DIKW Stages - COVID-19 (30 min)**

*(CSTA standards in this activity: 3A-DA-11, 3A-DA-12)*

- Students take ~10 minutes to review the DC COVID-19 dashboard from July 18, 2020: https://mayor.dc.gov/release/coronavirus-data-july-18-2020.
  - [Worksheet/Exit ticket] Students will answer: What do you recognize on this page? What did you learn? What are you left wondering about?
- Discuss graphs representing utilization of beds at acute care hospitals, cumulative incidence by Ward, and sustained decrease in cases within the community from July 18, 2020 (during the first summer lull in cases in the COVID-19 pandemic). What was the data? (cases) What information came from that data? (graphs) What knowledge can we learn? (can specify things like “Ward 4 has the highest cumulative number of cases”) What wisdom can we take away? (policy changes, etc.)
  - Some key features of this data source: time series data visualizations, map data visualization, and it’s a case where data science work gets the story “right” but only on a small time scale, because cases surged about two months after this.

**Understanding DIKW Stages - Heat Sensitivity Exposure Index (15 min)**

*(CSTA standards in this activity: 3A-DA-09, 3A-DA-10)*

- Students review the Heat Sensitivity Exposure Index dataset on OpenDataDC.
- Explain that Heat Sensitivity Exposure is a health issue in many cities worldwide.
- Discuss with the students:
  - What type of variable is the Total 2020 census population? [discrete]
  - What type of variable is the “Percent of population below 5 years of age”? [continuous]
- [Worksheet/Exit ticket] questions below:
  - What variables are stored in this dataset? Are they quantitative or qualitative variables?

This work was made possible through generous support from the National Science Foundation (Award # 2141655). 

- At what stage of analysis are we in the DIKW model?
- What would we need to do to progress through the later stages? What kinds of questions do you think we could answer with this dataset?

**Local Issue to Investigate (10 min)**

*(CSTA standards in this activity: 3A-AP-13, 3A-DA-12)*

- Students brainstorm local issues that they might be able to investigate with data.
- They talk to members of their small group, and write down some ideas to come back to in later lessons.

**Conclusion (5 min)**

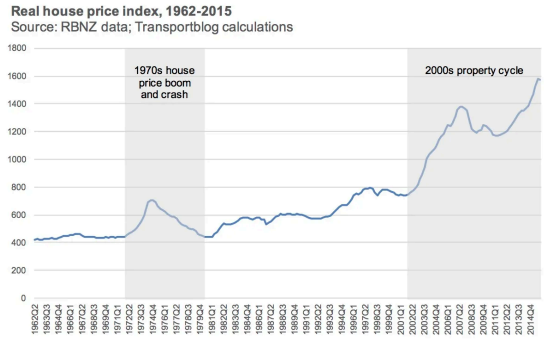
*(CSTA standards in this activity: 3A-DA-11, 3A-DA-12)*

- Review the key points of today's lesson.
  - What does each stage of the DIKW model look like?
  - Why do we need to analyze data instead of using it in its raw form?
  - Does the process of analysis always produce useful or correct wisdom? Why or why not?

This work was made possible through generous support from the National Science Foundation (Award # 2141655). 

**Exit Ticket (5 min)**

*(CSTA standards in this activity: 3A-DA-11, 3A-DA-12)*




- What stage of the DIKW model does the above represent?
- What do you think the cases or data in this investigation were?

**Assessment:**

Assess student understanding through participation in class discussions and class activities.

This work was made possible through generous support from the National Science Foundation (Award # 2141655). 