

Unit 2: Computational Foundations of Data Science

Lesson 2.1: What is Data Science?

In this lesson, students will talk about what data science is, experience a small-scale data science investigation, and watch a video for observations on all the activities that make up “data science.”

Duration: 90 minutes

Objective: By the end of this lesson, students will understand data science as an overlap of several fields, including programming, statistics, and mathematics. They will also have a rough understanding of the flow of a data science investigation.

Lesson Walkthrough Video: [Unit 2 Lesson 1 - Teacher Walkthrough](#)

CSTA Standards in this Lesson

Identifier	Concept	Subconcept	Standards
3A-DA-11	Data & Analysis	Collection Visualization & Transformation	Create interactive data visualizations using software tools to help others better understand real-world phenomena.
3A-DA-12	Data & Analysis	Inference & Models	Create computational models that represent the relationships among different elements of data collected from a phenomenon or process.
3A-IC-24	Impacts of Computing	Culture	Evaluate the ways computing impacts personal, ethical, social, economic, and cultural practices.
3A-IC-26	Impacts of Computing	Culture	Demonstrate ways a given algorithm applies to problems across disciplines.
3B-DA-06	Data & Analysis	Collection Visualization & Transformation	Select data collection tools and techniques to generate data sets that support a claim or communicate information.
3B-IC-25	Impacts of Computing	Culture	Evaluate computational artifacts to maximize their beneficial effects and minimize harmful effects on society.

Lesson activities

Class Discussion - Data Science (10 min)

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

- What do you think of when you think of data science?
- Highlight variety of responses: data visualization, big data management, programming, mathematical modification of data, etc.

Food Deserts in DC Exploration (30 min)

(CSTA standards in this activity: 3A-DA-11, 3A-DA-12, 3A-IC-24, 3A-IC-25)

- The class works together on the question, “Where are the food deserts in DC?”
- Students read the article [Food Deserts in DC](#) to understand the problem and consider further information they want to explore.
- Students use the OpenDataDC dataset [Grocery Store Locations](#) to identify the number of grocery stores and locations by ward.
- Using locations of grocery stores from the OpenDataDC dataset, students identify the closest Metro stop to each grocery store, and calculate distances.
- Using any tool they want to use, or a [new data table in CODAP](#), students create a graph to model information to answer their question.
- What more information would we want to continue this investigation?

TikTok Viral Video Tracking (30 min)

(CSTA standards in this activity: 3A-DA-12, 3A-IC-24, 3B-IC-25)

- Students watch and discuss “[We tracked what happens after TikTok songs go viral](#)” and talk about the techniques and things they saw.
- What data do you expect them to collect?
- At 9:30 into the video, discuss:
 - Why are record companies trying to sign artists with viral TikTok hits?
 - How do you think they know who to sign?
 - What artist is going viral this week?
 - NEXT SLIDE:
 - Spotify releases “editorial playlists” put out by the company. What effect do you think this has on record companies’ control on the music market?
 - Does this create a better or worse situation for independent artists?
- Note: this video shows several different elements/stages of a data science investigation, including data processing and data visualization.



Lecture - Data Science (10 min)

(CSTA standards in this activity: 3A-IC-26, 3A-DA-12)

- Data science is the overlap of programming, statistics, machine learning/AI, social science applications/applications in the field using DS
 - Data scientists choose questions and problems they want to explore.
 - They determine appropriate data to collect, reason about quantitative and categorical data, rules of probability, and use appropriate computational tools to analyze that data.
 - They use summary statistics, tables, and data visualizations to present their findings and share conclusions.
 - This process normally spawns more questions to investigate!
- (The data science investigation cycle is also shown on the next slide.)

Exit Ticket (10 min)

(CSTA standards in this activity: 3B-DA-06)

- Find an interesting dataset within OpenDataDC. Paste the link to the dataset in the first response.
- What does this dataset include? (variable types? what kind of source is it? etc.)
- Why do you think this dataset is important/interesting?
- What is one question you have about this data?

Assessment:

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

