# Unit 4: Data Collections
## Introduction to Using Data Sets

Aligned to:

AP Computer Science A Course and Exam Description

Effective Fall 2025

Topic 4.2



https://longbaonguyen.github.io

# What Is a Data Set?

- A data set is a collection of specific pieces of information.
- Test scores for a class. Daily rainfall for a year. Every transaction a shop made last month.
- What makes it a data set rather than a pile of numbers is that the pieces are related and can be examined together to answer a question.

# How We Work With One

- You almost never look at a data set all at once. You work through it.
- Values in the set are accessed and used ONE AT A TIME, and each one is processed toward whatever outcome you are after.
- That pattern (visit each value, do something with it, carry a result forward) is called a traversal, and the rest of this unit builds on it.
- Arrays, ArrayLists and files are three ways to hold a data set, and the traversal is much the same in each.

# Plan Before You Code

- Data can be represented in a chart or a table. That picture is how you plan the algorithm.
- Laying the data out shows you what one row is, which column you care about, and what you are accumulating as you go.
- Describe the algorithm in plain language first. If you cannot state it in English, writing Java will not make it clearer.

# A Data Set to Work With

- A cafeteria waste audit, five school days:

| Day | Meals served | Pounds discarded |
|---|---|---|
| 1 | 412 | 38 |
| 2 | 398 | 51 |
| 3 | 405 | 33 |
| 4 | 421 | 47 |
| 5 | 388 | 62 |

- Each row is one day. Each column is one thing measured about that day.

# From Table to Algorithm

- Question: which day wasted the most food per meal served?
- In words, before any Java:
- 1) Start with no best day yet.
- 2) For each row, compute pounds discarded divided by meals served.
- 3) If that is bigger than the best so far, remember this day instead.
- 4) When the rows run out, the remembered day is the answer.
- That description is the algorithm. Writing it in Java is then a matter of translating each step.

# Three Questions

- When you are staring at a data set and do not know where to start, answer these:
- 1) What am I looking at each step?  (one row? one number?)
- 2) What am I carrying with me?  (a running total? a count? the best so far?)
- 3) What do I do when the data runs out?  (return it? divide it? print it?)
- Almost every algorithm in this unit is those three answers put into a loop.

# Algorithms to Know

- Over a data set, you will be asked to:
- total the values
- count the ones that meet a condition
- find the largest or smallest
- compute an average
- find the item with the biggest something (not the biggest value itself)
- check whether every value, or any value, satisfies a test
- Once you know the shape of these, most questions become a variation on one of them.

# Your Turn

- Using the cafeteria table, describe each of these in words, including what you carry, what you compare, and how you finish:
- 1) What was the average number of meals served per day?
- 2) How many days had more than 45 pounds of waste?
- 3) Which day had the best ratio of meals served to pounds discarded?
- Do not write Java yet. Write sentences, then compare with a partner. Any place you disagree is a place the question was ambiguous.

# References

1) Building Java Programs: A Back to Basics Approach by Stuart Reges and Marty Stepp

2) Runestone CSAwesome Curriculum: https://runestone.academy/runestone/books/published/csawesome/index.html

For more tutorials/lecture notes in Java, Python, game programming, artificial intelligence with neural networks:

https://longbaonguyen.github.io