Data Collections

Theresa Migler-VonDollen
CMPS 5P
Many programs deal with large collections of similar information.
  - Words in a document
  - Students in a course
  - Data from an experiment

Recall the programming assignment where we asked the user to input grades and we returned the average.
  - That program didn’t keep track of the actual numbers, just the sum.
  - What if we also wanted to compute the median?
The median is the data value that splits the data into equal-sized parts.

For the data 1, 4, 5, 9, 42, the median is 5, since there are two values greater than 5 and two values that are smaller.

One way to determine the median is to store all the numbers, sort them, and identify the middle value.
We need a way to store and manipulate an entire collection of numbers.

Can we just use a lot of variables?
  - No, because we don’t know how many we will need at the start.

We need some way of combining an entire collection of values into one object.
Recall that Python lists are ordered sequences of items.

A list or array is a sequence of items where the entire sequence is referred to by a single name and individual items can be selected by indexing.

In other programming languages, arrays are generally a fixed size, meaning that when you create the array, you have to specify how many items it can hold.

Arrays are generally also *homogeneous*, meaning they can hold only one data type.
Python lists are dynamic. They can grow and shrink on demand.

Python lists are also heterogeneous, a single list can hold arbitrary data types.

Python lists are mutable sequences of arbitrary objects.
Aside from all of the list operations that we have already seen, there is also the membership operation.

3 in ourList

Practice with:

- ourList.reverse()
- ourList.sort()
- ourList.count(2)
- ourList.insert(5,'Thanks!')
- ourList.remove(7)
Most of these methods don’t return a value - they change the contents of the list in some way.

Lists can grow by appending new items, and shrink when items are deleted. Individual items or entire slices can be removed from a list using the `del` operator.
Write a Python program that takes a list of numbers and calculates the mean and median.

- If the list has odd length, the middle value in the list is the median.
- If the list has even length, the median is the average of the middle two values.
All of the list examples we’ve looked at so far have involved simple data types like numbers and strings.

We can also use lists to store more complex data types.
Write a Python function that takes a list and removes all duplicate values from the list.