* priority queue

want A.insert(key, value)
A.retrieveMax()

fast

* what if

want A.insert(key, value)
A.retrieve(key)

fast ?

map $key_1 \rightarrow value_1$
$key_2 \rightarrow value_2$
⋮
$key_K \rightarrow value_K$

unordered / ordered

dictionary $key_1 \rightarrow value_{1.1}$, $value_{1.2}$, ⋮

		search insert	retrieve
* sparse array (or dense) linked list	w/ key	O(1)	O(n)
	w/ ordered key	O(n)	O(log n)
	w/ key	O(1)	O(n)
	w/ ordered key	O(n)	O(n)

* if keys are integers
O(1) insert O(1) retrieve
by dense array (that wastes space)

* how about using a "simple" (almost O(1)) function
h: key → integer

A.insert(key, value) ⇒ A[h(key)] = value

v = A.retrieve(key) ⇒ v = A[h(key)]

bucket array — hash function

* if {keys} ⟺ (one-to-one) {0, 1, 2, ⋯, k-1}

"perfect hashing"

* what if not ?

e.g. # keys > k

⇒ two keys of the same h(·)

⇒ collision !

* e.g. h(str) = str[0] - 'a' ; 26 possibilities

h("act") = h("apple")

$h(str) = \sum_i (str[i] - 'a')$ 2^{32} ? possibilities not uniform

$h(str) = \sum_i (str[i] - 'a') \% k$ k possibilities

"stop" "pots" "spot"

⇒ easy collision

* how to hash

h(key) = compress (Hashcode(key))

- Hashcode(key):
  - sum [not using position info]
  - polynomial (like shift-left)
  - cyclic shift
- compress:
  - division (% k)
  - MAD ((a·□+b) % k)

* how to resolve collision during insertion ?
  ① fixed (unordered) array per bucket
     [can still overflow]
  ② linked list per bucket
     chaining : don't want long chains
  ③ other data structure [usually called secondary] per bucket
     [more complicated]
  ④ use other empty buckets
     open addressing

* open addressing
  a[key] = value
  x = a[key]

  ① insert (key, value) to check $h_0(key) = h(key)$
  ② if fail, insert to check $h_1(key)$
  ③ if fail, insert to check $h_2(key)$
  ⋮
  (m-1) if fail, insert to check $h_m(key)$
  (m) declare failure
     not found

* (A) linear probing :

$$h_i(key) = (h_{i-1}(key) + 1) \% K$$
$$= (h_0(key) + i) \% K$$

primary clustering

$m = K-1$

(B) quadratic probing :

$$h_i(key) = (h_0(key) + i^2) \% K$$

secondary clustering

$m = ?$

(C) double hashing :

$$h_i(key) = (h_0(key) + i\,\tilde{h}(key)) \% K$$

$\tilde{h}(key) > 0$

$m = ?$