Office Checklist

- Get speakers
- Turn on talking hat
- Tape it into hat
- Get iphone5

Preclass Checklist

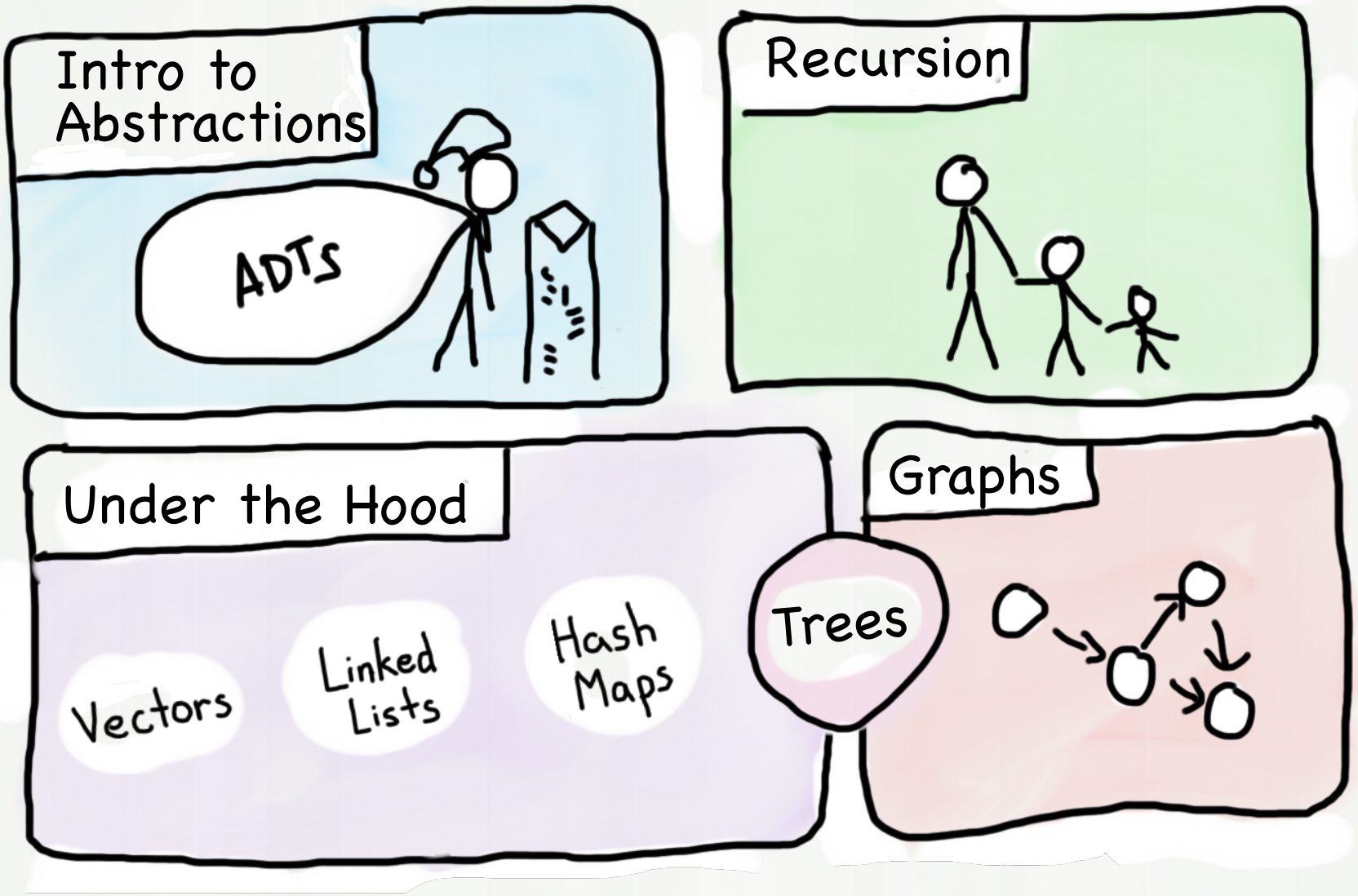
- Play music
- Tape down seat locations
- Put up/out handouts
- Set up bins
- Get sorting hat accessible
- Turn on sorting hat speakers and test
- Turn on the projector
- Run HashMap code (and make sure caffeine is on)
- Test a few Shazam runs
- Test out the presentation pointer
- Make sure all code is loaded into QT and starter code is empty
- Running: QT, Powerpoint, Ink2Go

Announcement: PQ

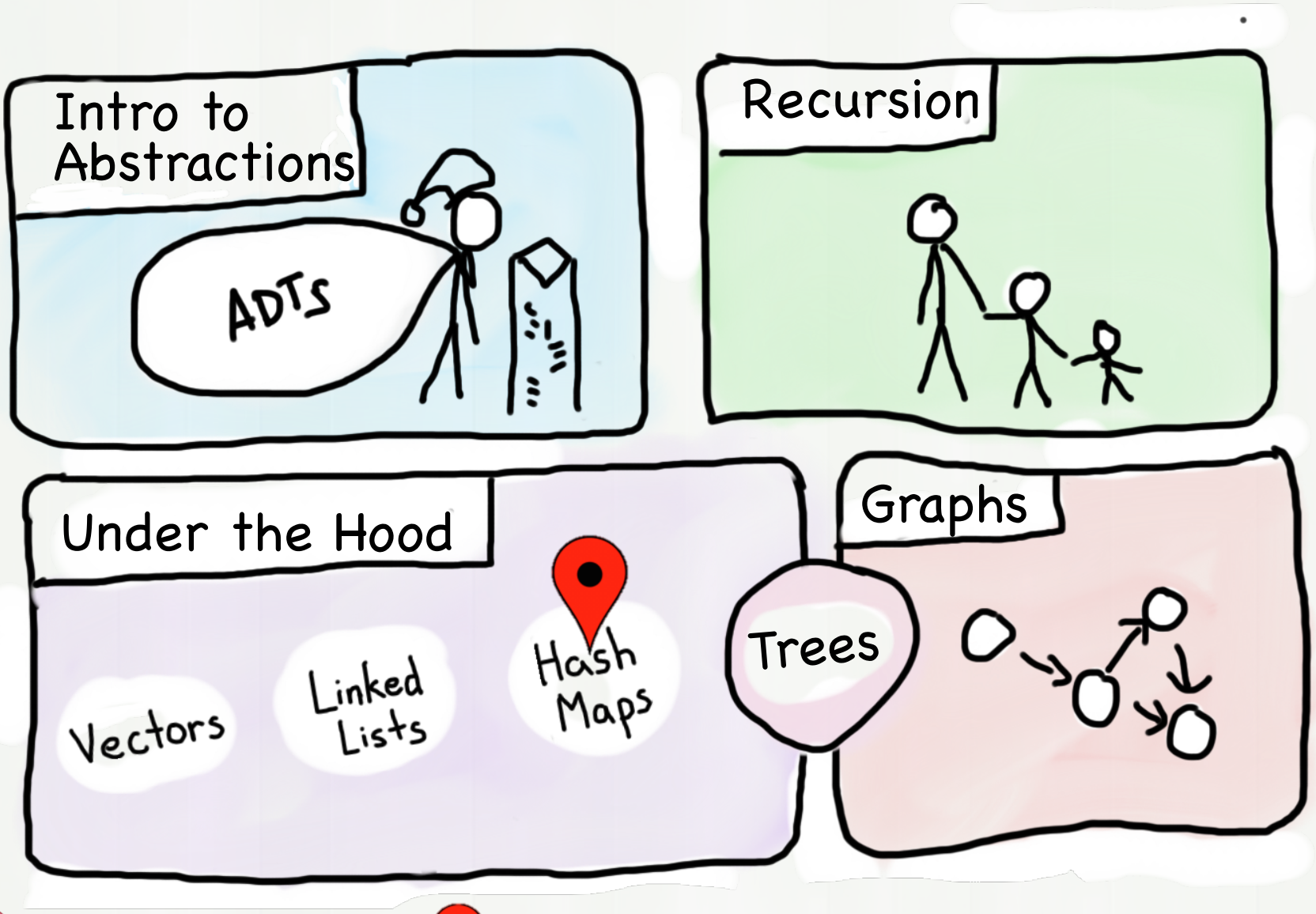
Start
Feb 15nd

Due
Feb 24th

Programming Abstractions



Programming Abstractions



First, a cool program

Shazam

Source: Shazam

By the end of today I am going to show you how this works.

We've Gotten Ahead of Ourselves

Source: The Hobbit

Start at the Beginning

CS106B

Chris Piech

Hashing: The Heart of the Hash Map

Today's Goals

1. Understand how the HashMap works
2. Add hashing to your algorithmic toolbox
3. Practice linked lists

Today's Route

(eg HashMap<string, string>)

Other Applications

Code

Real Hash fns

Collisions

Hashing

HashMaps River

First Objective

(eg HashMap<string, string>)

Wikipedia API

```
// adds name / text pair to dataset
put(string articleName, string articleHTML)

// returns corresponding articleHTML
get(string articleName)

// removes the article
remove(string articleName)
```

Wikipedia is big!

5 million articles and growing

Source: xkcd

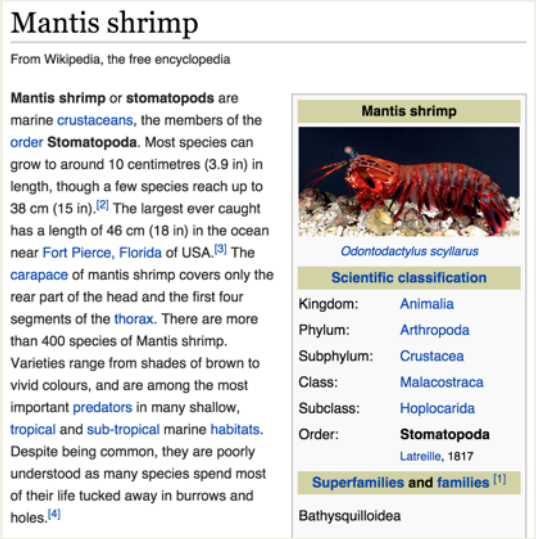
Key Value Pairs

Key = Title, Value = Article

Key:

"Mantis Shrimp"

Value:



Mantis shrimp

From Wikipedia, the free encyclopedia

Mantis shrimp or **stomatopods** are marine crustaceans, the members of the order **Stomatopoda**. Most species can grow to around 10 centimetres (3.9 in) in length, though a few species reach up to 38 cm (15 in).[2] The largest ever caught has a length of 46 cm (18 in) in the ocean near Fort Pierce, Florida of USA.[3] The carapace of mantis shrimp covers only the rear part of the head and the first four segments of the thorax. There are more than 400 species of Mantis shrimp. Varieties range from shades of brown to vivid colours, and are among the most important predators in many shallow, tropical and sub-tropical marine habitats. Despite being common, they are poorly understood as many species spend most of their life tucked away in burrows and holes.[4]

Mantis shrimp

Odontodactylus scyllarus

Scientific classification	
Kingdom:	Animalia
Phylum:	Arthropoda
Subphylum:	Crustacea
Class:	Malacostraca
Subclass:	Hoplocarida
Order:	**Stomatopoda** Latreille, 1817

Superfamilies and families [1]

Bathysquilloidea

Key:

"Antelope Canyon"

Value:

Antelope Canyon is a slot canyon in the American Southwest. It is located on Navajo land east of Page, Arizona. Antelope Canyon includes two separate, photogenic slot canyon sections, referred to individually as *Upper Antelope Canyon* or *The Crack*; and *Antelope Canyon* or *The Corkscrew*.[2]

The Navajo name for Upper Antelope Canyon is Tsé bighánílíní, which means "the place where water runs through rocks." Lower Antelope Canyon is Hazdistazí (advertised as *"Hasdestwazi"* by the Navajo Parks and Recreation Department), or "spiral rock arches." Both are located within the LeChee Chapter of the Navajo Nation.[4]

Contents [hide]

1 Geology
2 Tourism and photography
2.1 Upper Antelope Canyon

Antelope Canyon

Tsé bighánílíní dóó Hazdistazí (Navajo)

A beam of light in Upper Antelope Canyon

Antelope Canyon

Key Value Pairs

Key = Title, Value = Article

Key:

"Mantis Shrimp"

Value:

```
"<p><b>Mantis shrimp</b>
or <b>stomatopods</b> are
marine <a href="/wiki/
Crustacean"
title="Crustacean">crusta
ceans</a>, the members of
the <a href="/wiki/
Order_(biology)"
title="Order
(biology)">order</a>
<b>Stomatopoda</b>.
though a few species
reach up to 38 ..."
```

Key:

"Antelope Canyon"

Value:

Antelope Canyon is a slot canyon in the American Southwest. It is located on Navajo land east of Page, Arizona. Antelope Canyon includes two separate, photogenic slot canyon sections, referred to individually as *Upper Antelope Canyon* or *The Crack*; and *Antelope Canyon* or *The Corkscrew*.[2]

The Navajo name for Upper Antelope Canyon is Tsé bighánílíní, which means "the place where water runs through rocks." Lower Antelope Canyon is Hazdistazí (advertised as *"Hasdestwazi"* by the Navajo Parks and Recreation Department), or "spiral rock arches." Both are located within the LeChee Chapter of the Navajo Nation.[4]

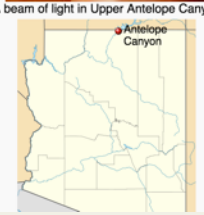
Contents [hide]

1 Geology
2 Tourism and photography
2.1 Upper Antelope Canyon

Antelope Canyon

Tsé bighánílíní dóó Hazdistazí (Navajo)

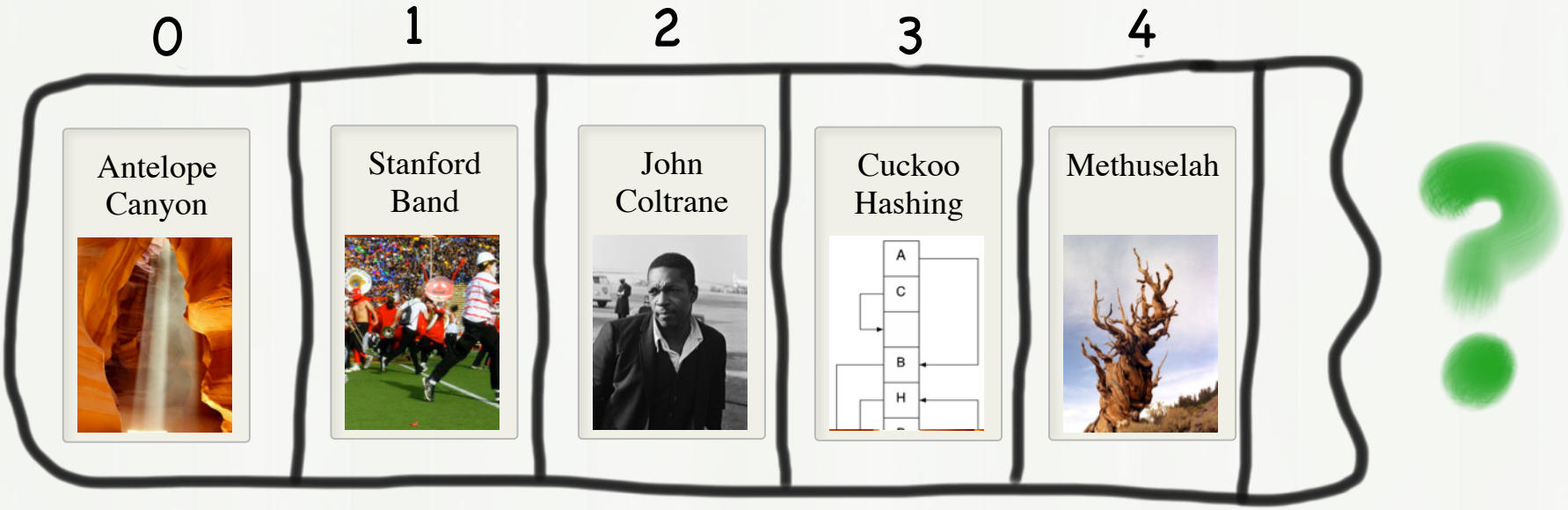
A beam of light in Upper Antelope Canyon



Thought Experiment

What if: we store articles in one gigantic vector

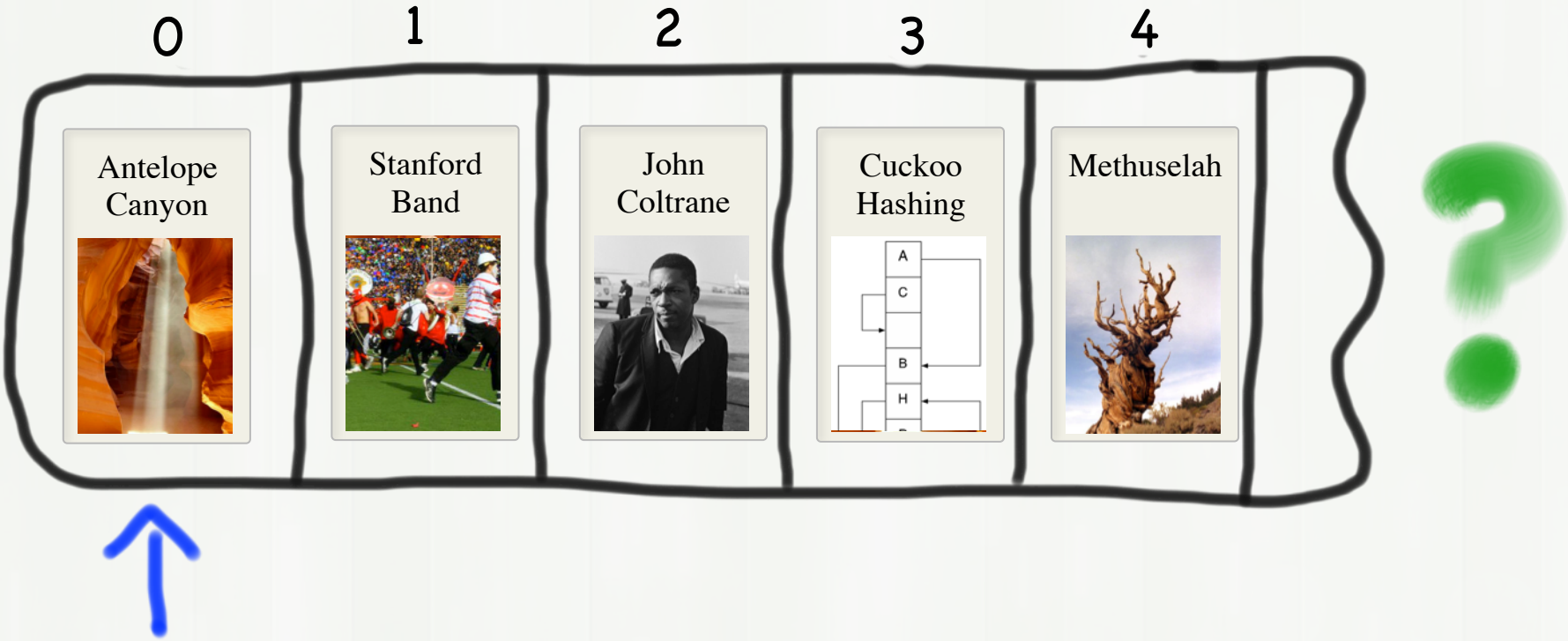
Wikipedia:



Thought Experiment

What if: we want to `get("Mantis shrimp")`?

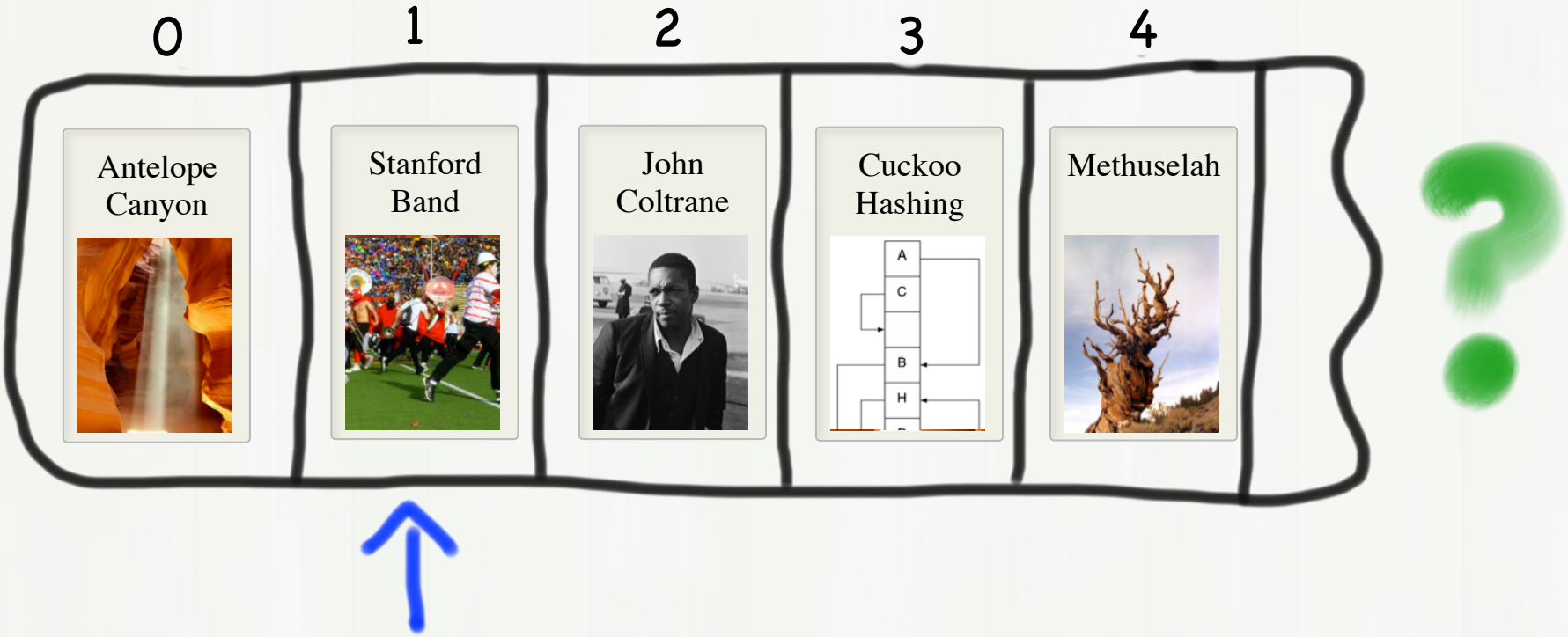
Wikipedia:



Thought Experiment

What if: we want to `get("Mantis shrimp")`?

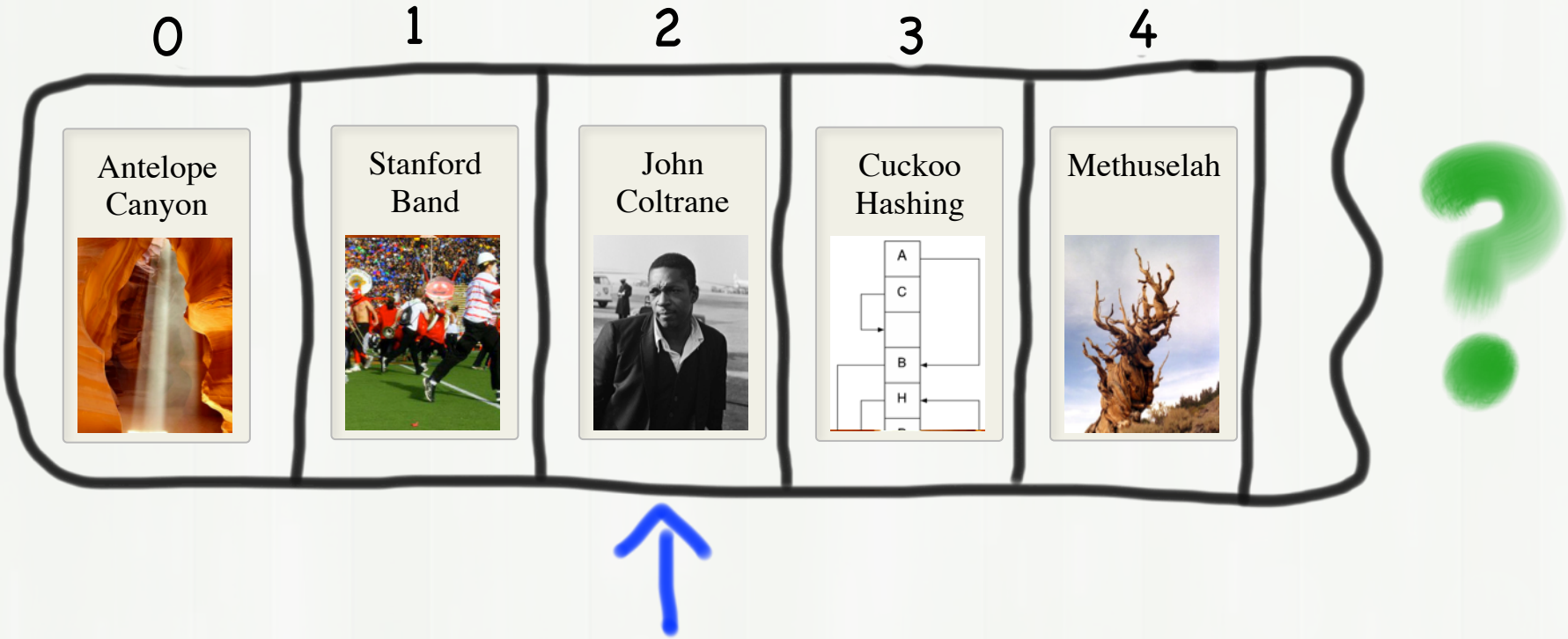
Wikipedia:



Thought Experiment

What if: we want to `get("Mantis shrimp")`?

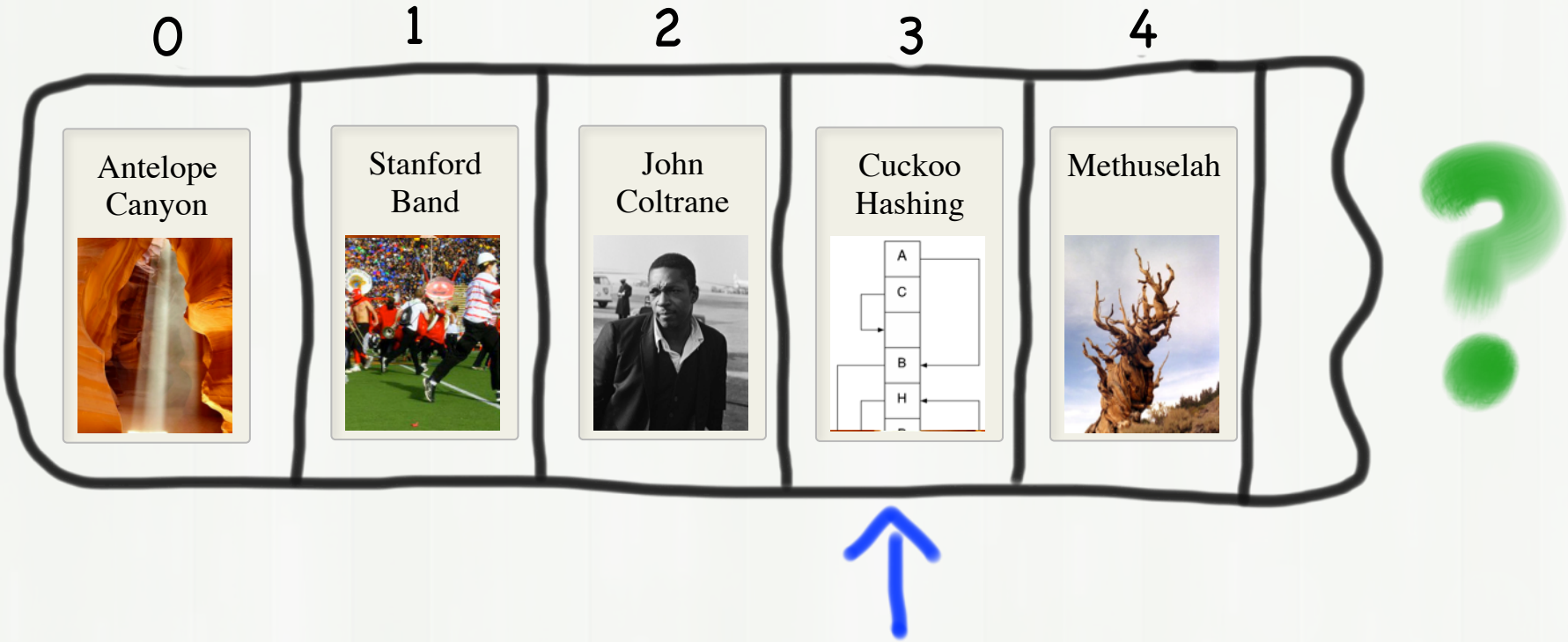
Wikipedia:



Thought Experiment

What if: we want to `get("Mantis shrimp")`?

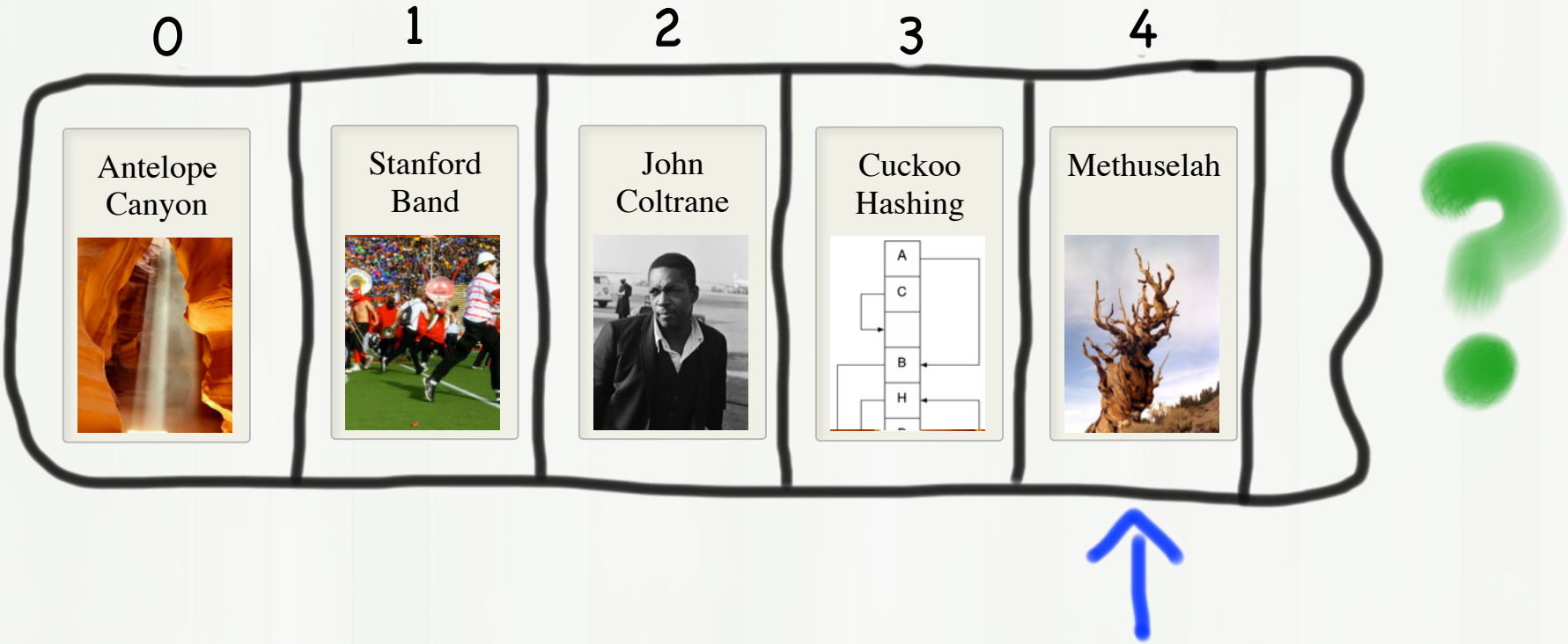
Wikipedia:



Thought Experiment

What if: we want to `get("Mantis shrimp")`?

Wikipedia:



Has to be a better way!

What if: there was a way to know where our article belonged just by looking at the title?

Hash Function

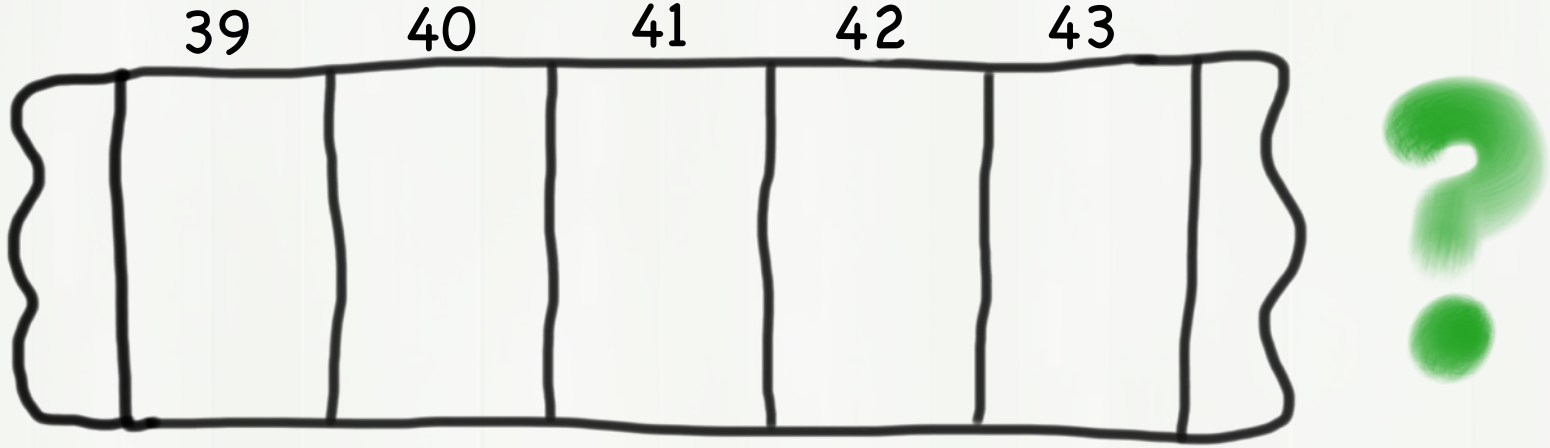

```
int hash(string key);
```

Hash Function


```
int hash(string title);
```

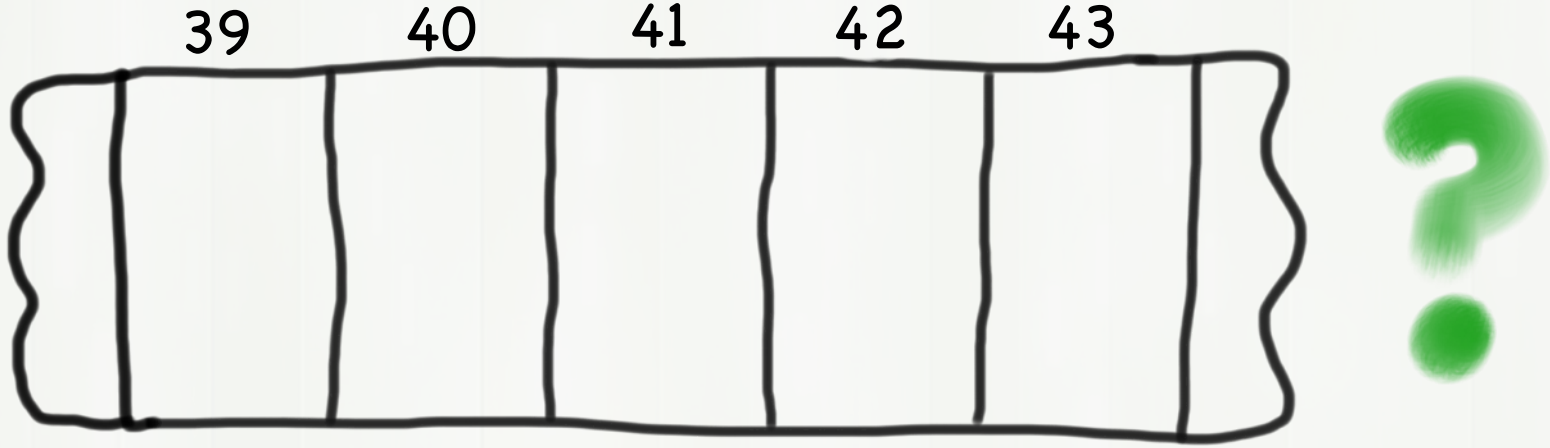
Wikipedia with Hashing

Wikipedia:



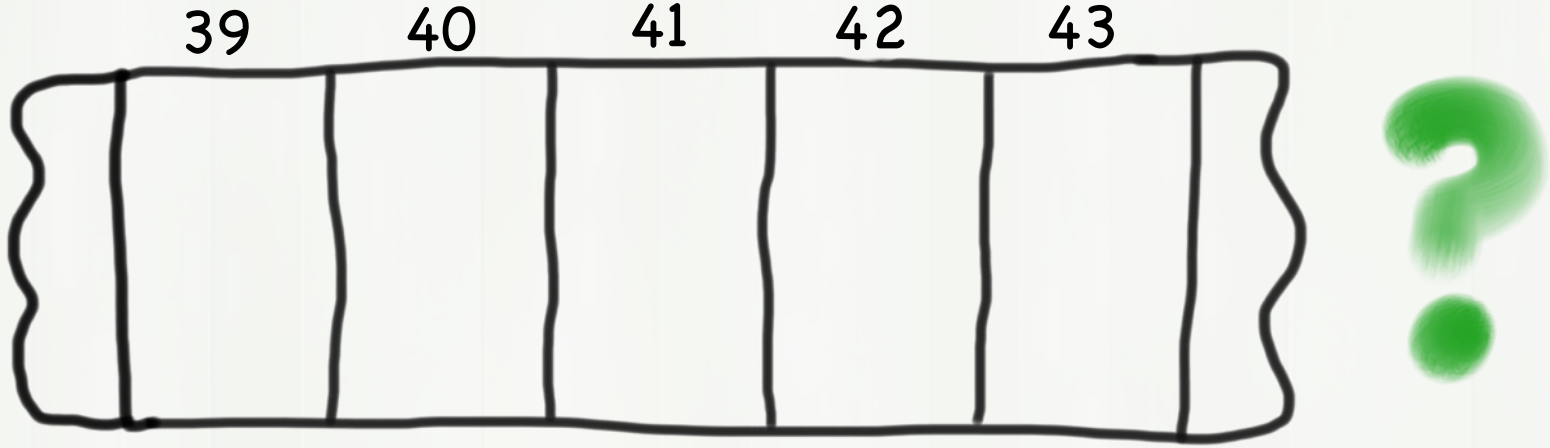
put("mantis shrimp", html)

Wikipedia:



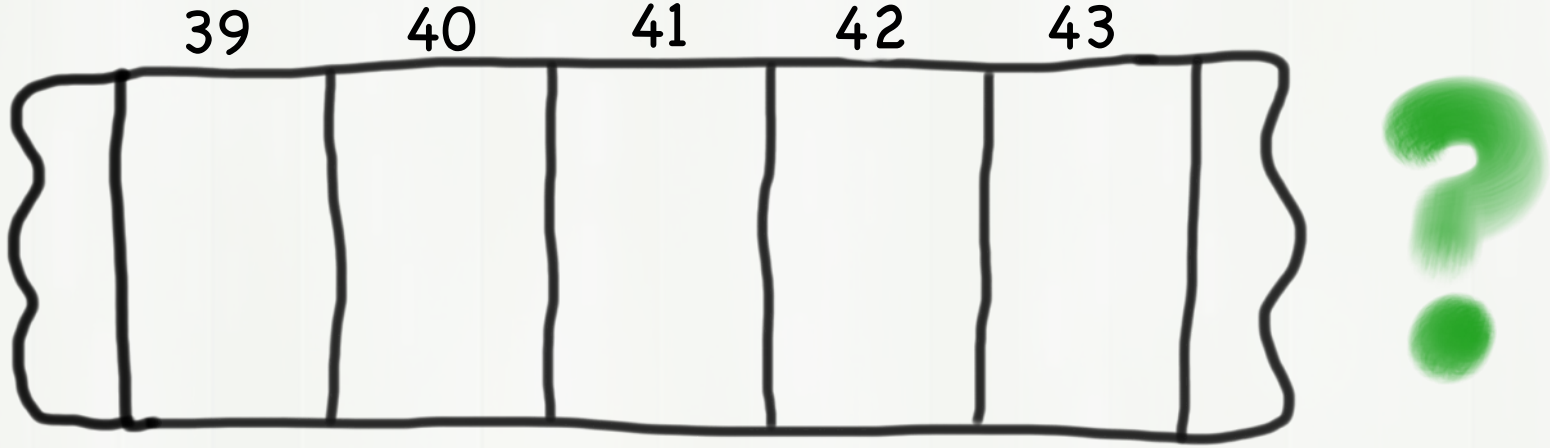
put("mantis shrimp", html)

Wikipedia:



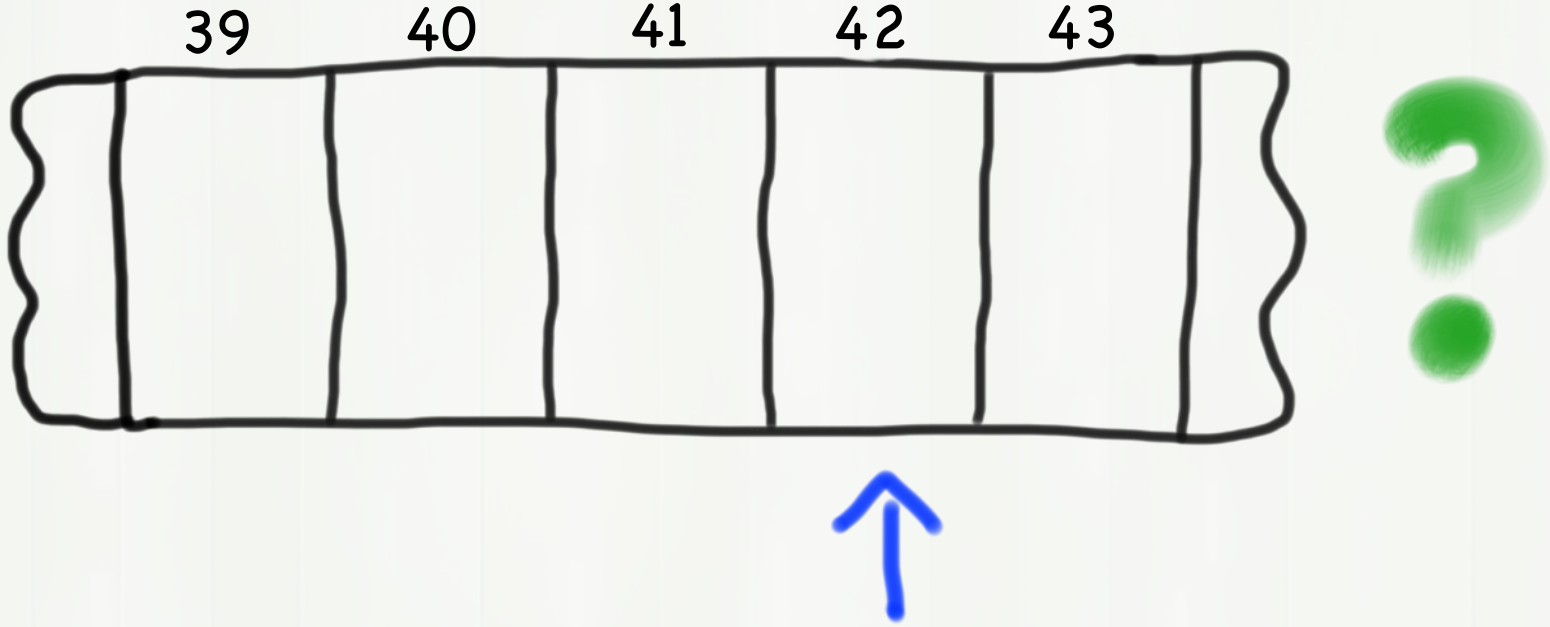
put("mantis shrimp", html)

Wikipedia:



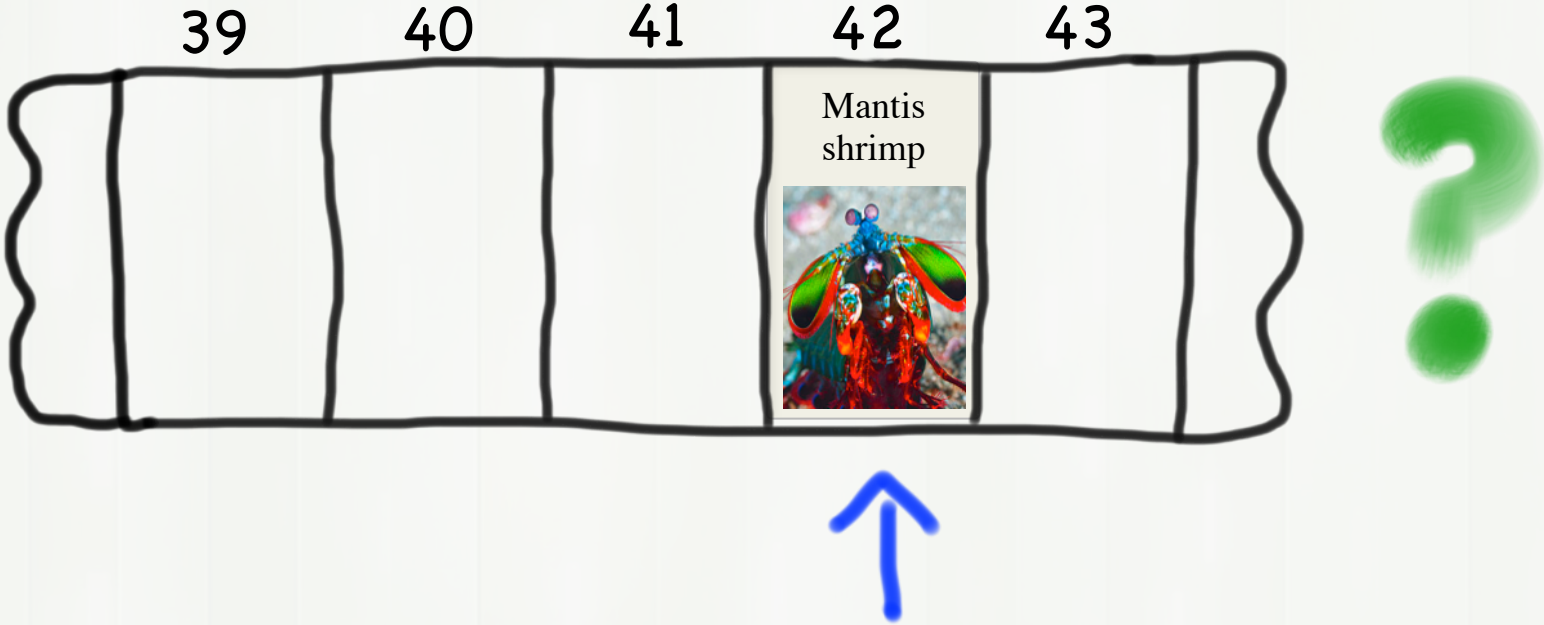
put("mantis shrimp", html)

Wikipedia:



put("mantis shrimp", html)

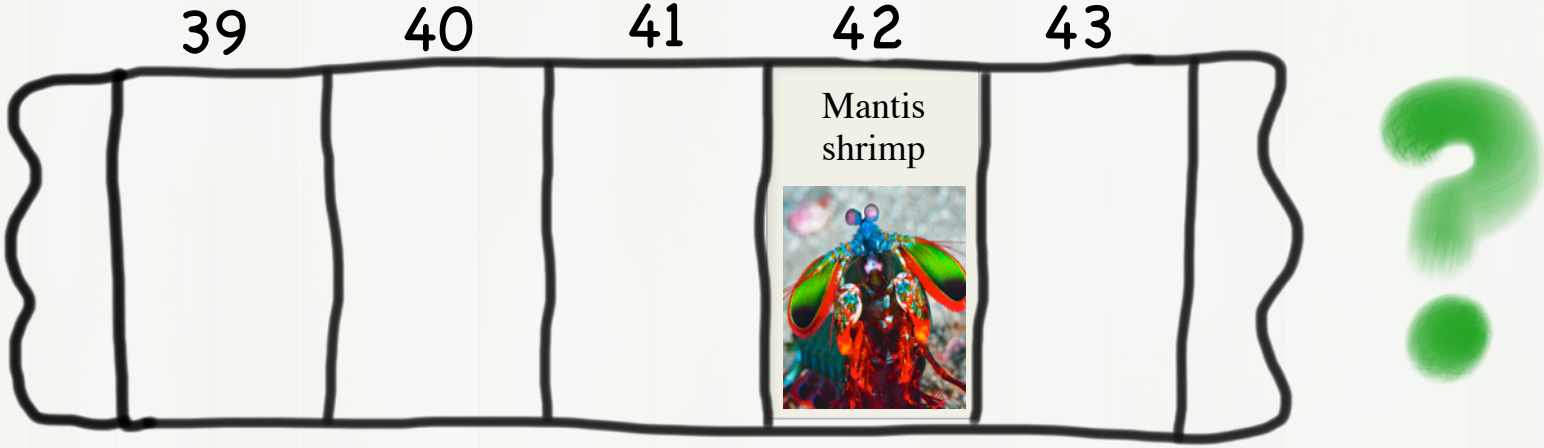
Wikipedia:



Wonderful!

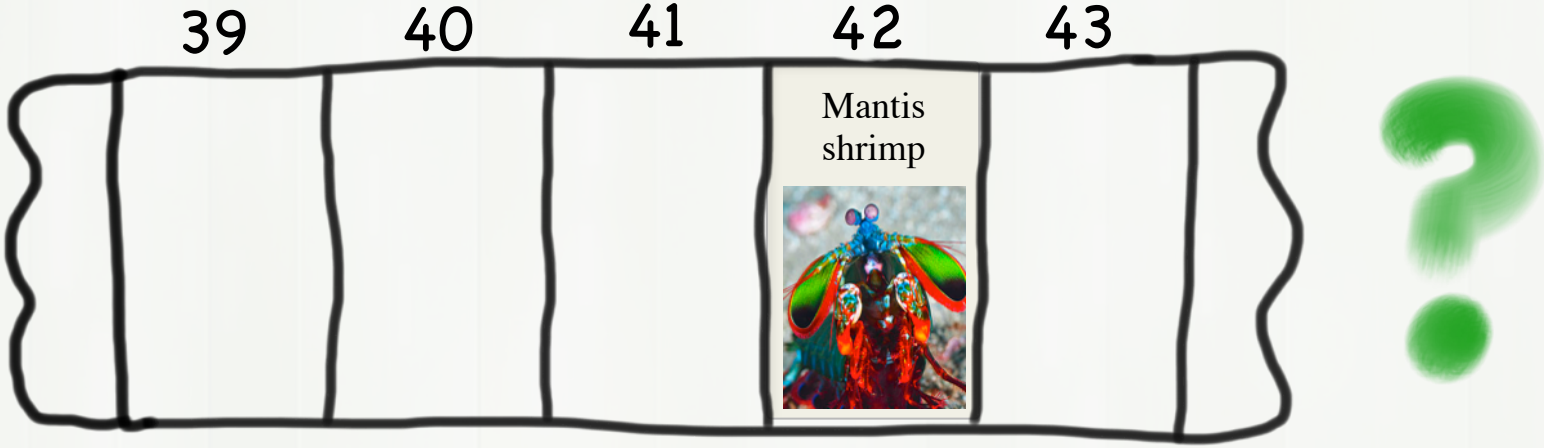
get("mantis shrimp")

Wikipedia:



get("mantis shrimp")

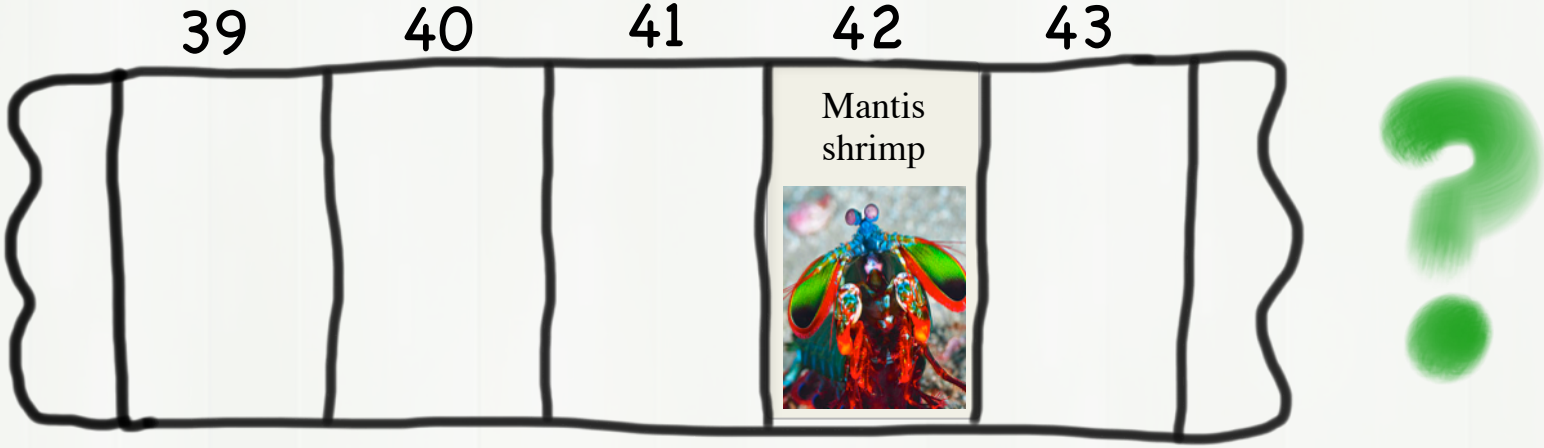
Wikipedia:



"mantis shrimp" → Hash Fn

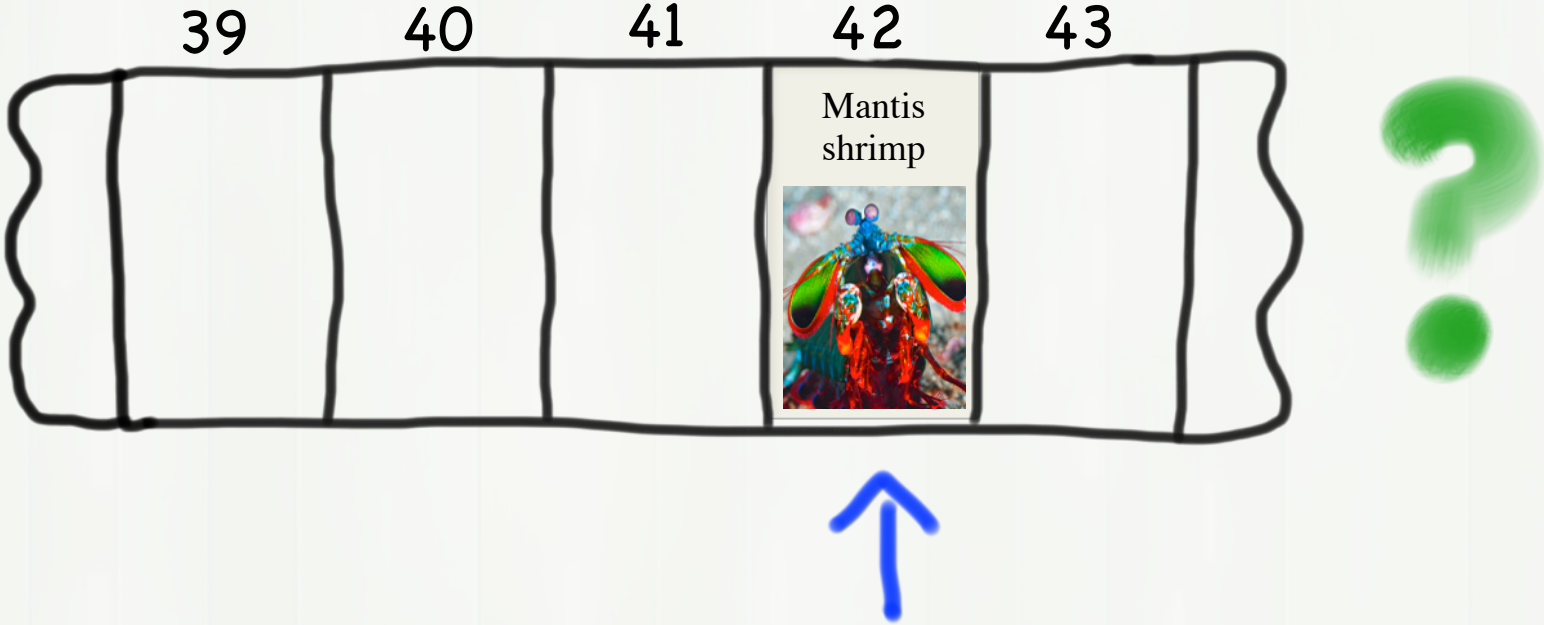
get("mantis shrimp")

Wikipedia:



get("mantis shrimp")

Wikipedia:



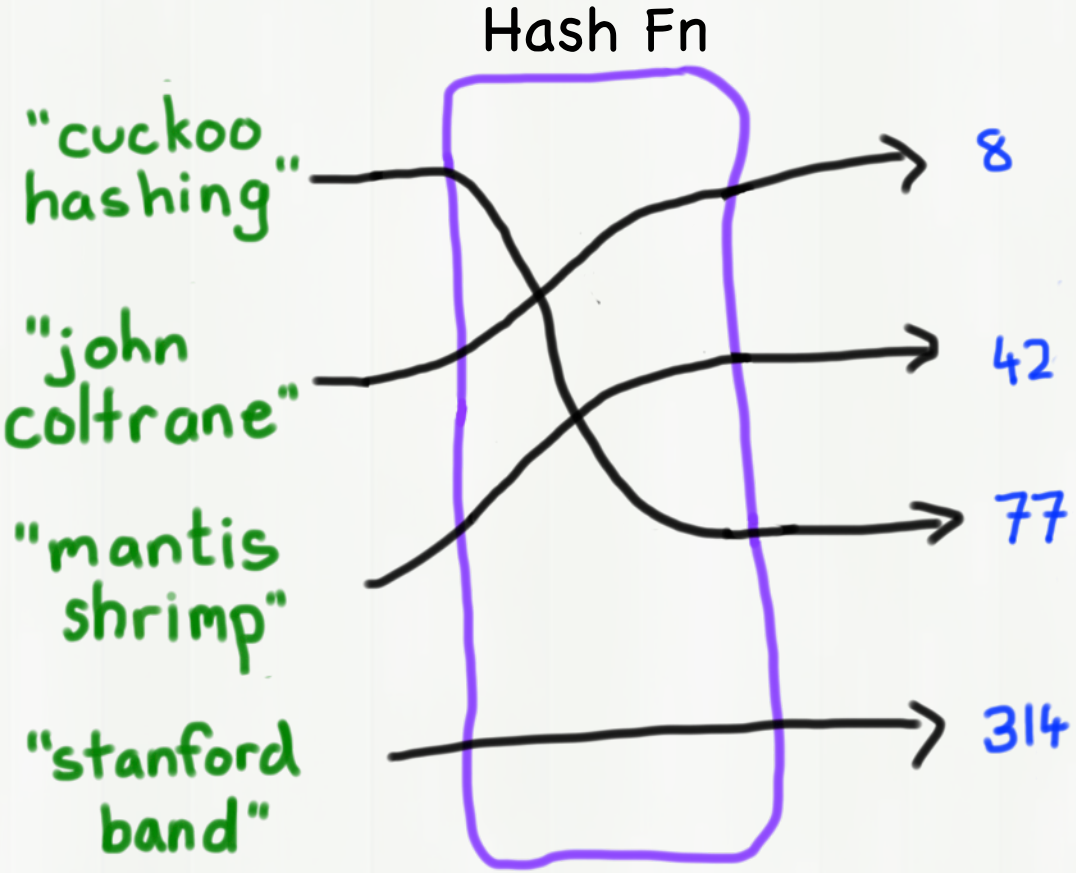
Source: Planet Animal Zone

Property #1: Consistent

If you pass in the same input, you will ***always*** get the same index.

Property #2: Well Distributed

If you pass in the different inputs, a hash function will return different indices as often as possible.



Does one exist?

What Does a Hash Fn Look Like?

You have already seen one!

```
int hash(string key) {
   int name0 = alpha(key[0]);
   int name1 = alpha(key[1])'
   int preHash = name0 + 7 * name1;
   return preHash % NUM_BUCKETS;
}
```

Volunteer

Today's Route

(eg HashMap<string, string>)

Other Applications

Code Wikipedia

Real Hash fns

Collisions

Hashing

HashMaps River

Today's Route

(eg HashMap<string, string>)

Other Applications

Code Wikipedia

Real Hash fns

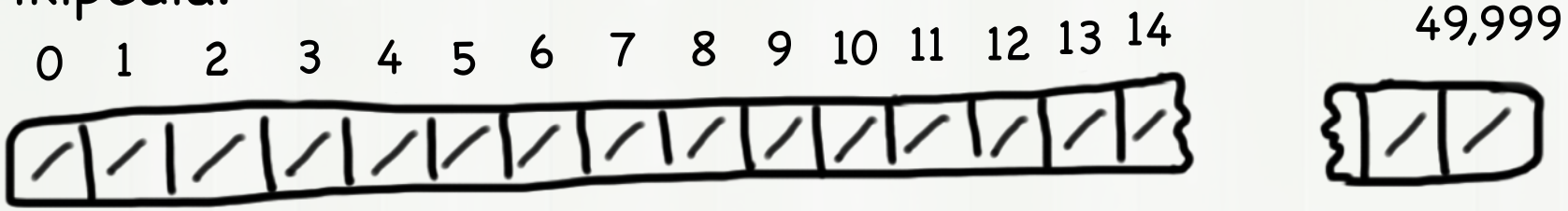
Collisions

Hashing

HashMaps River

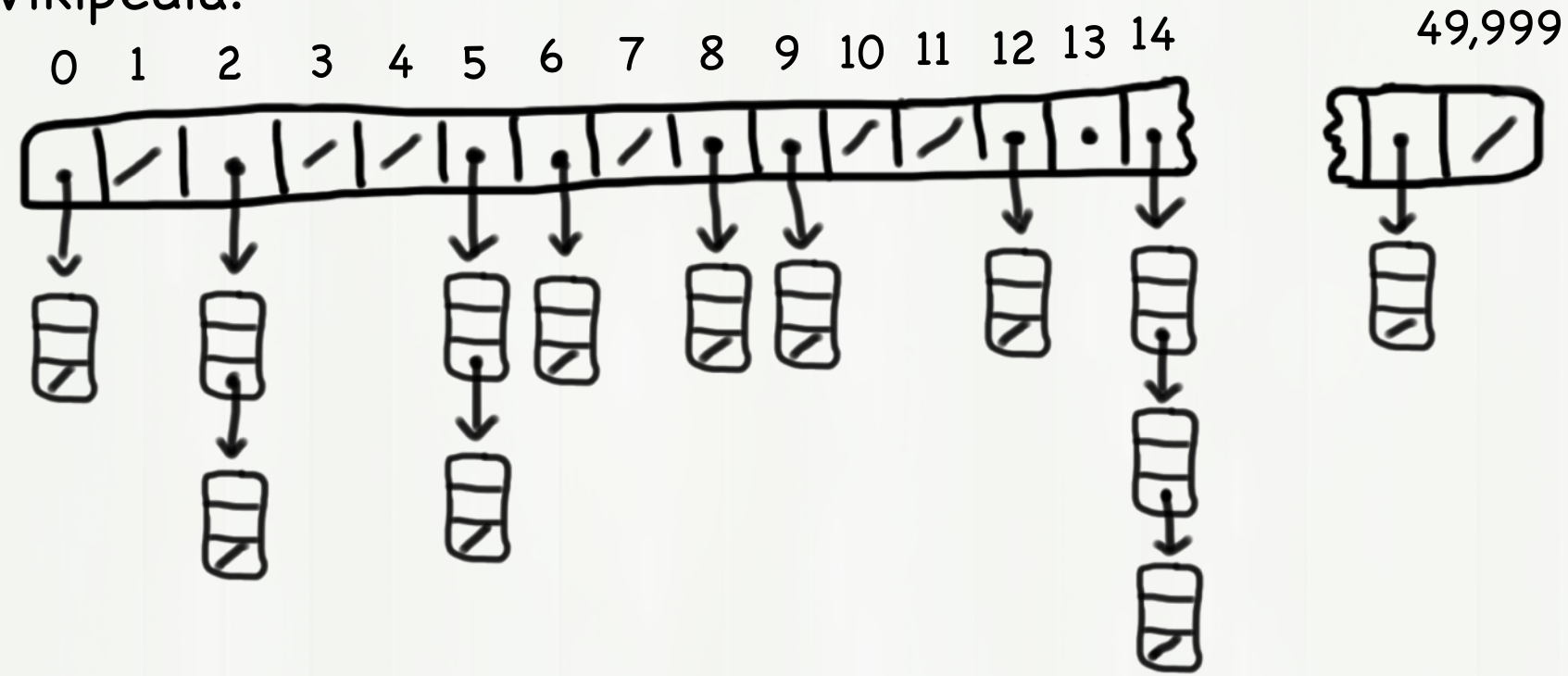
HashMap

Wikipedia:



HashMap

Wikipedia:

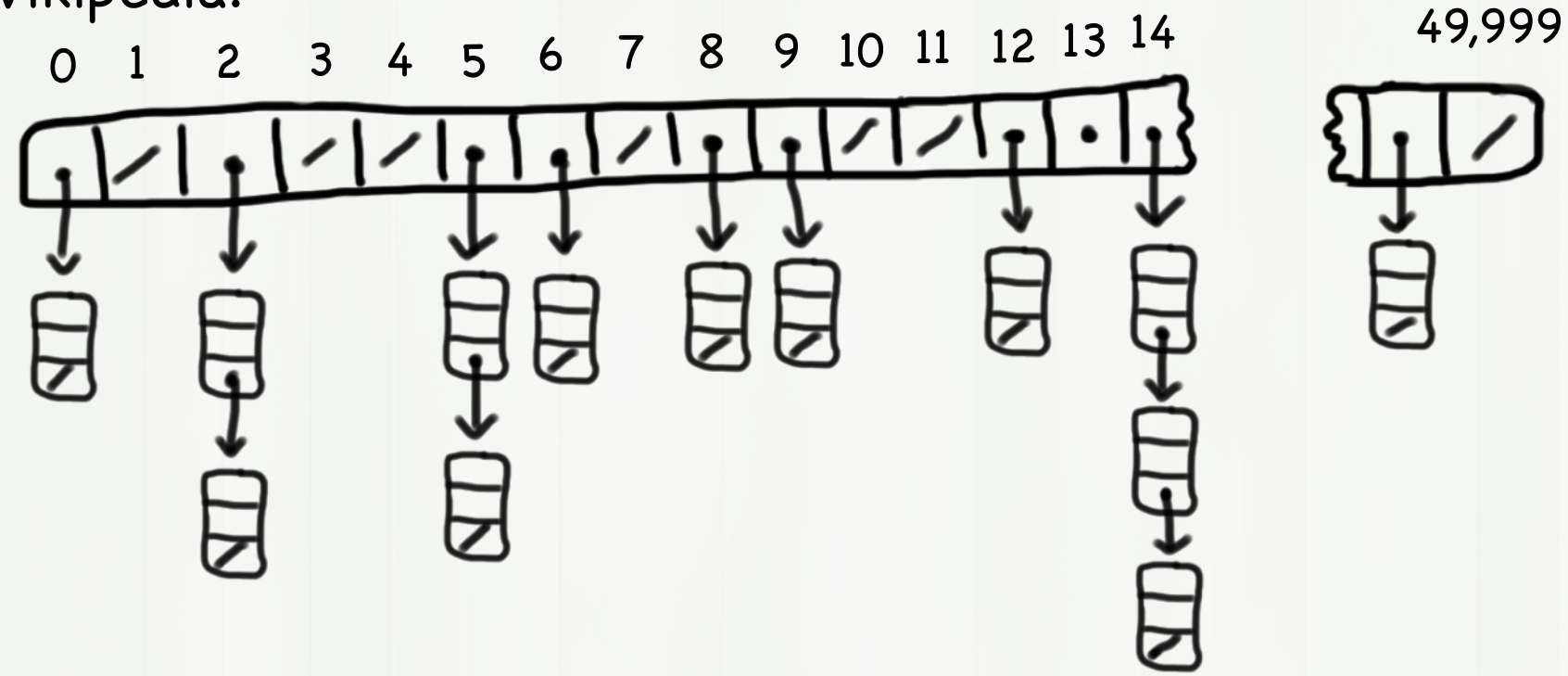


Hash Fn

Our old friend, linked lists!

HashMap

Wikipedia:



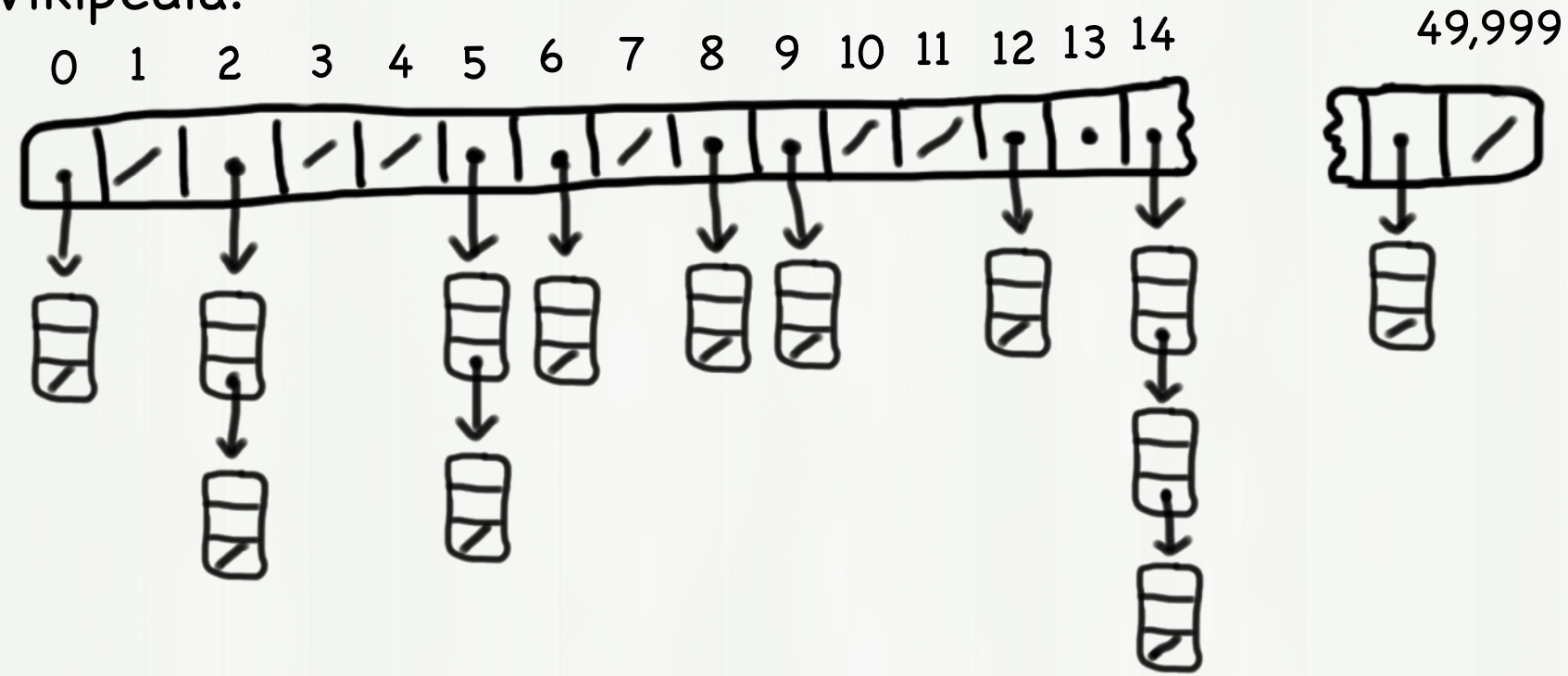
Article:

Key
Value
Next

Hash Fn

HashMap

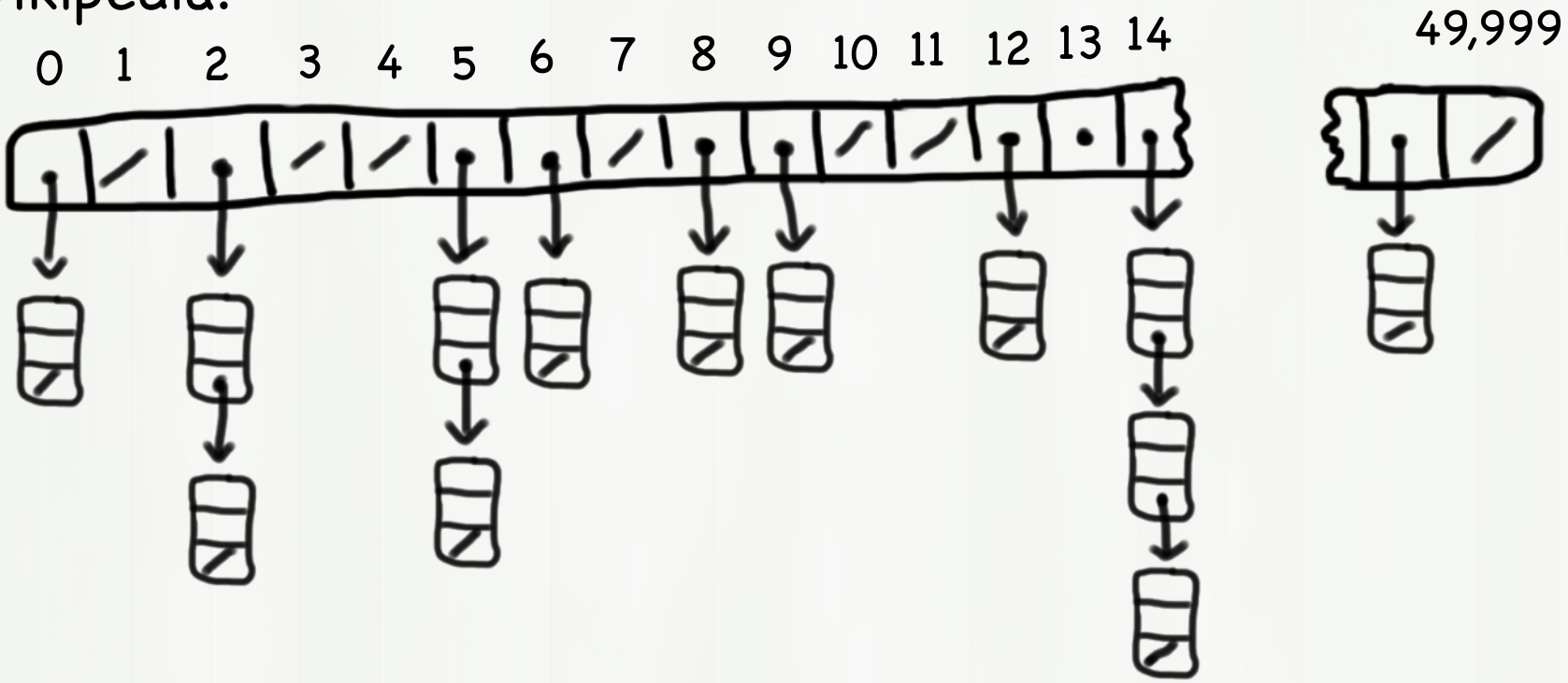
Wikipedia:



Hash Fn

get("Antelope Canyon")

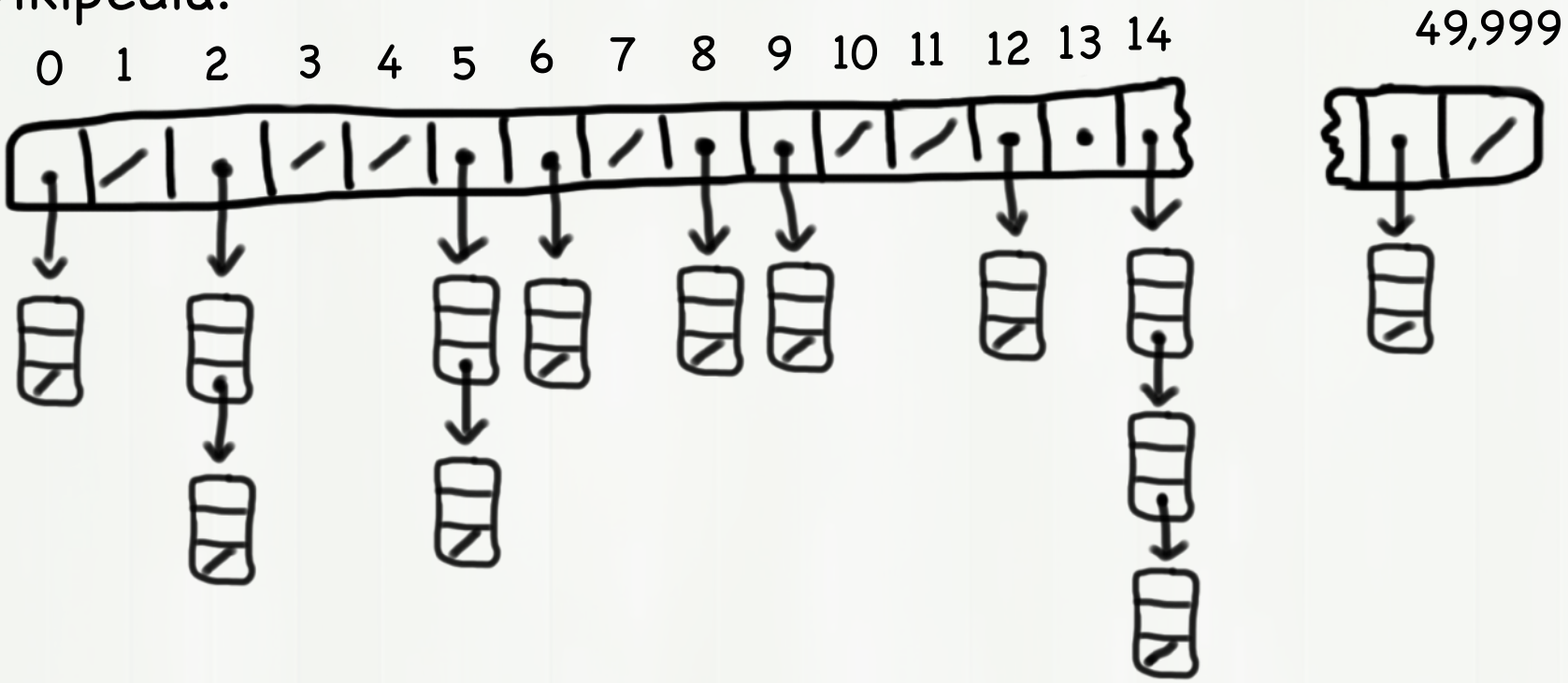
Wikipedia:

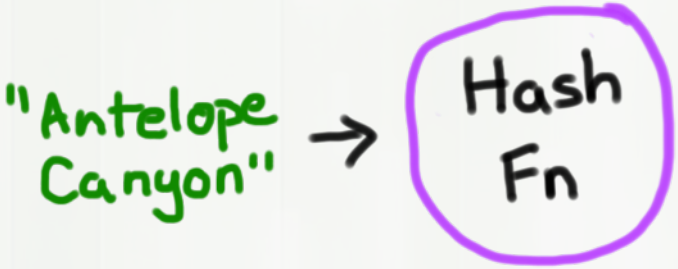


Hash Fn

get("Antelope Canyon")

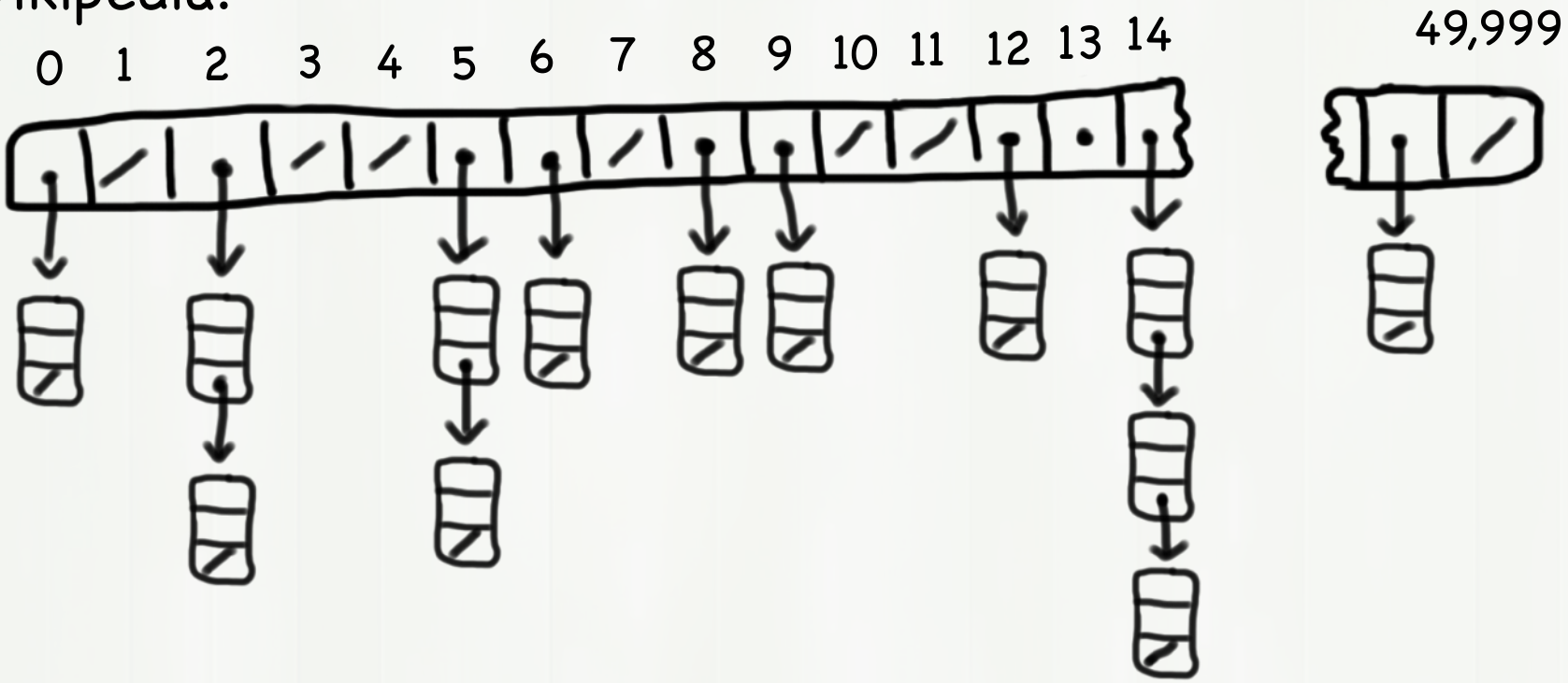
Wikipedia:



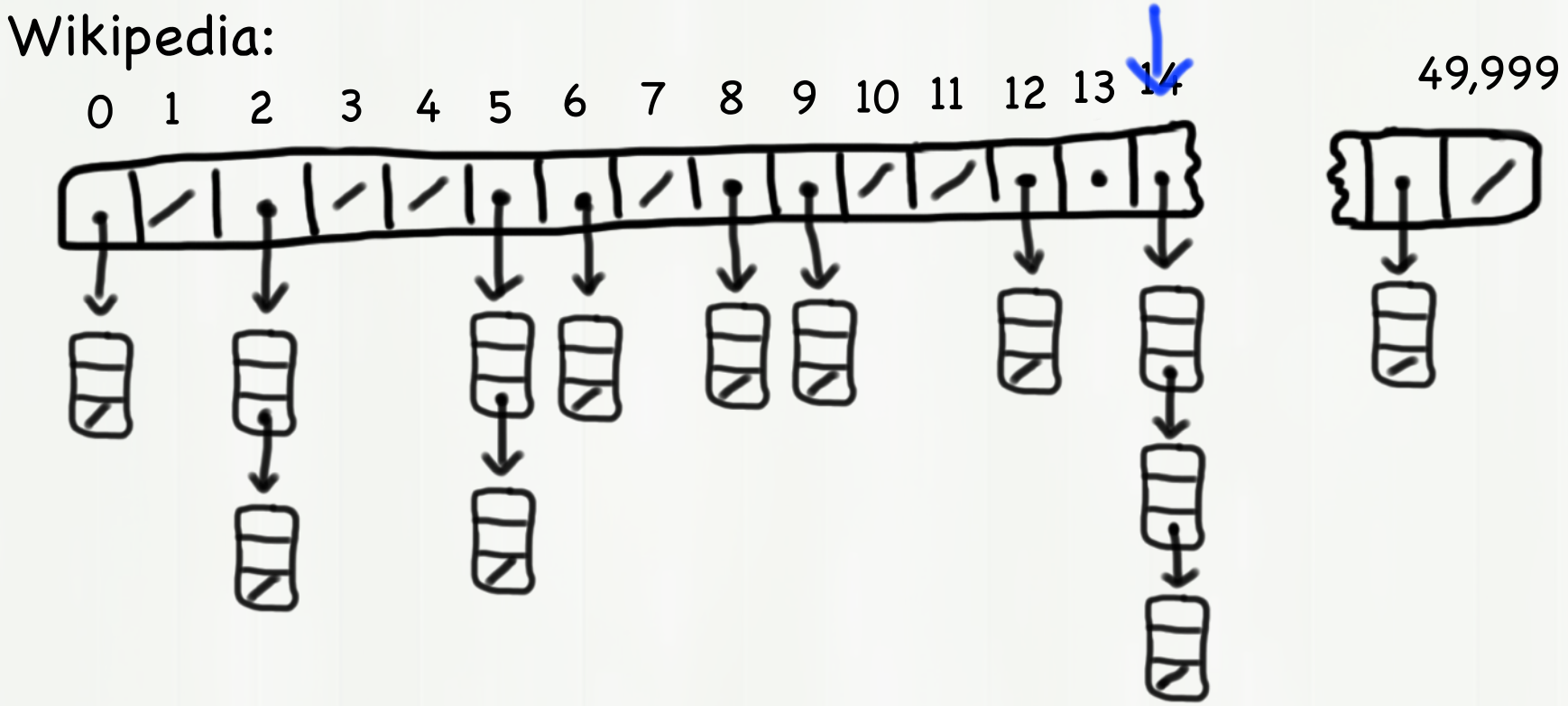


get("Antelope Canyon")

Wikipedia:

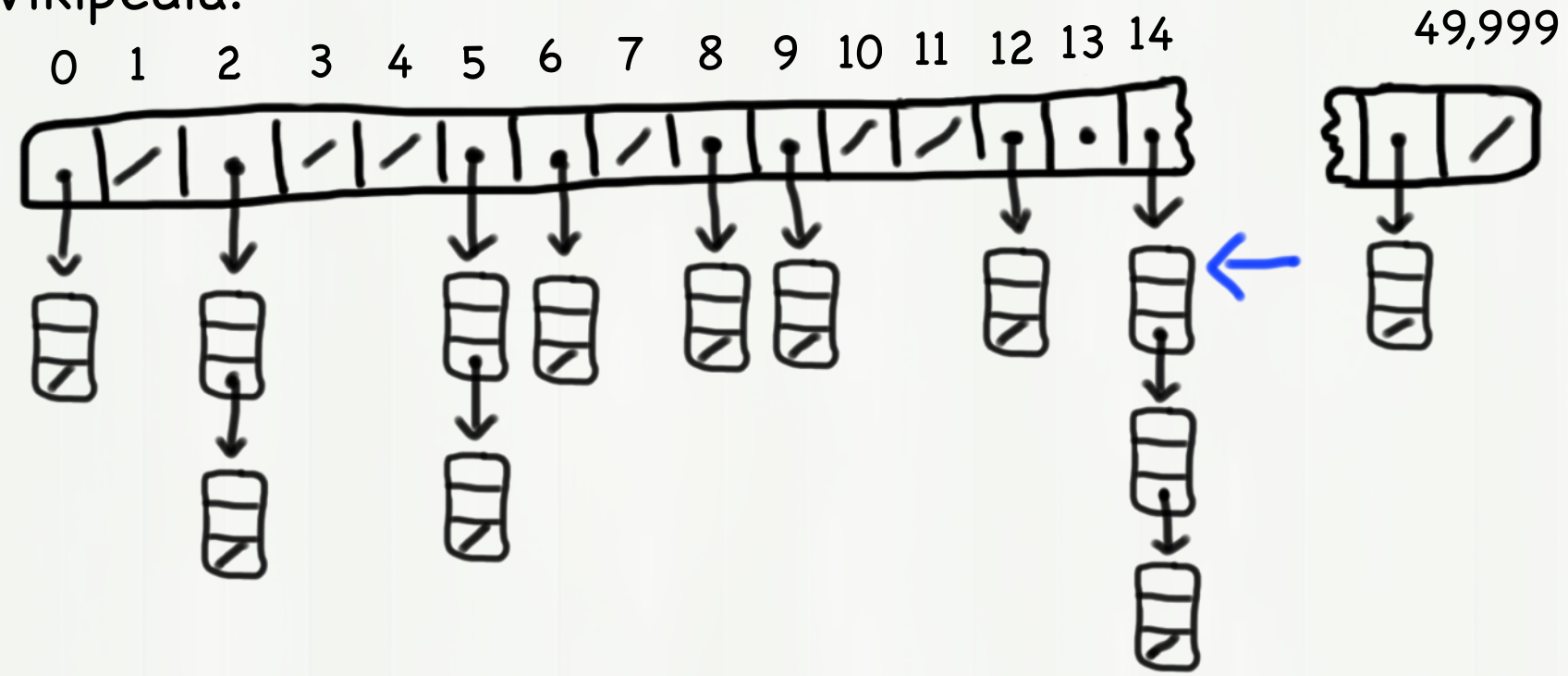


get("Antelope Canyon")



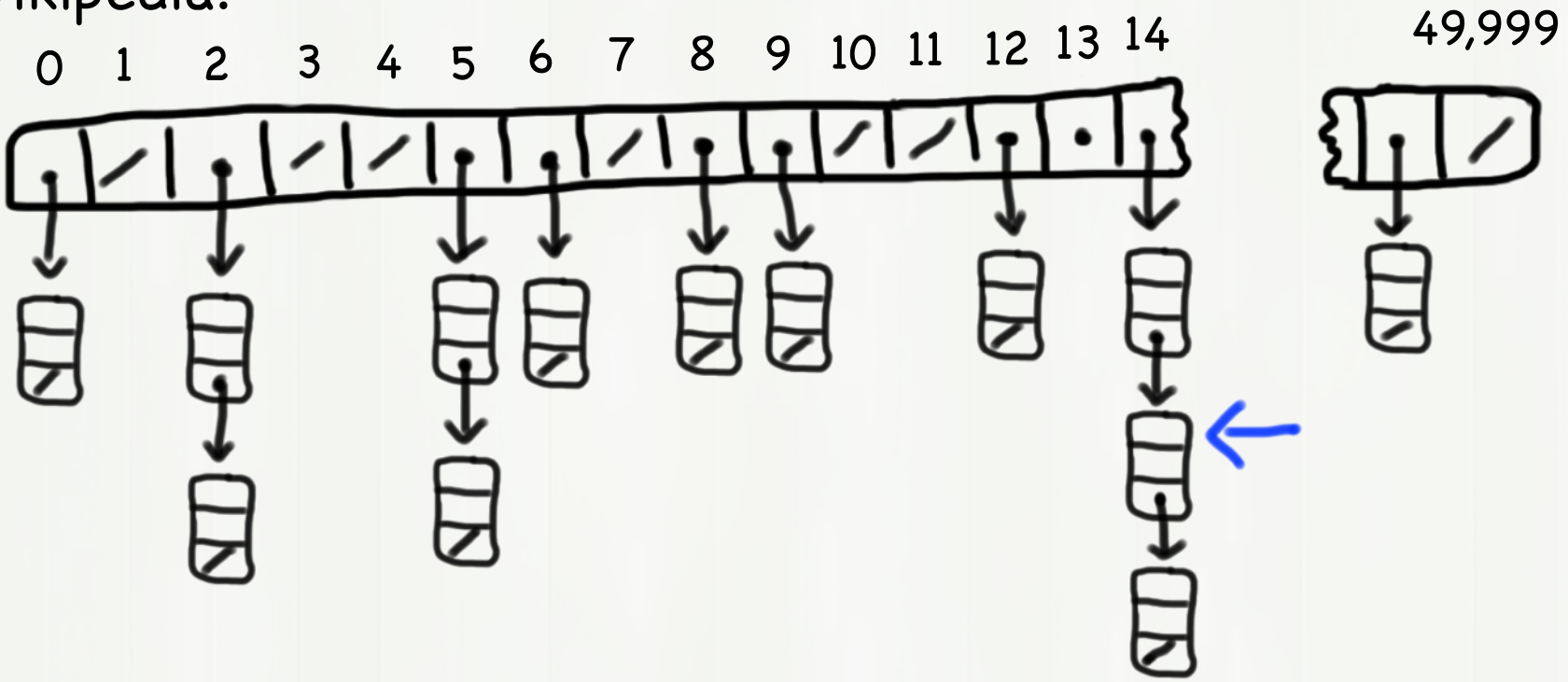
get("Antelope Canyon")

Wikipedia:



get("Antelope Canyon")

Wikipedia:



get("Antelope Canyon")

Wikipedia:

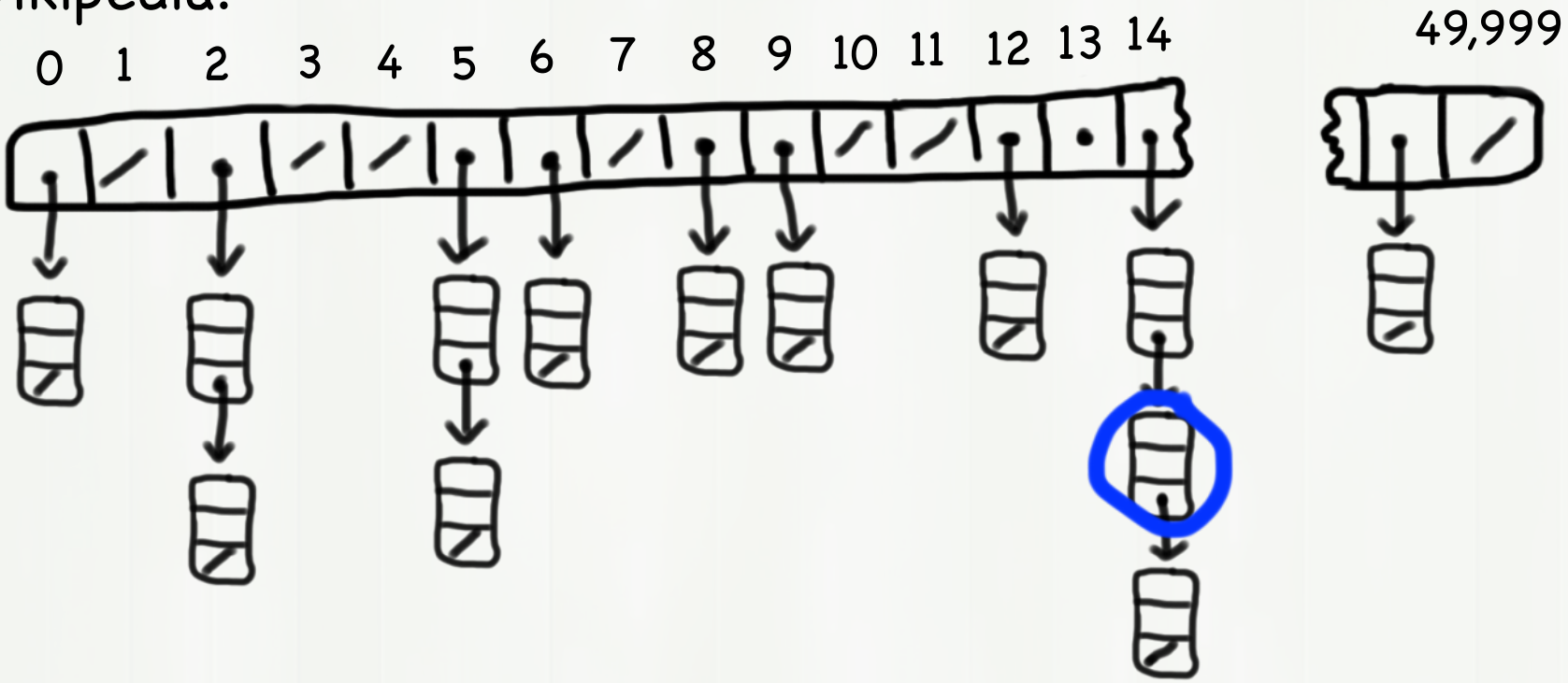
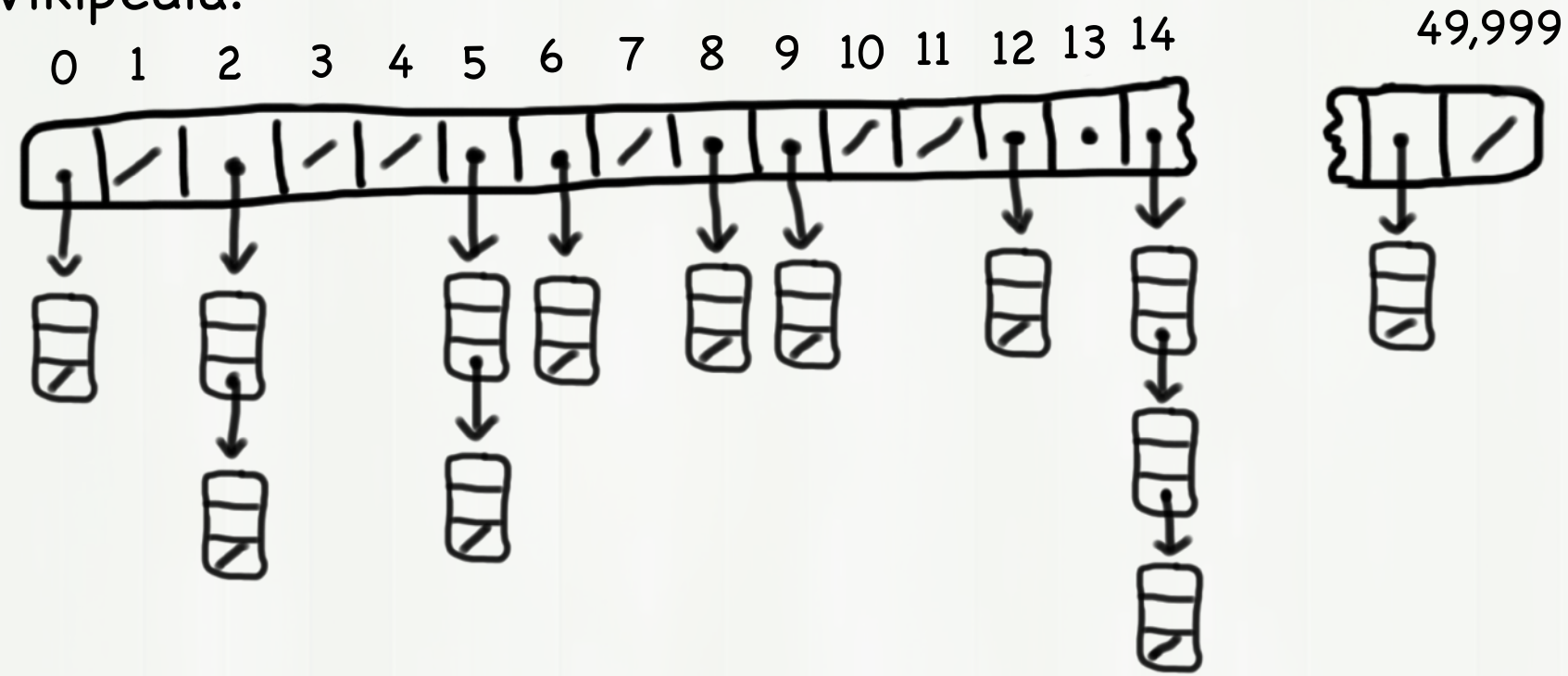


Photo Credit: Alex Mironyuk

HashMap

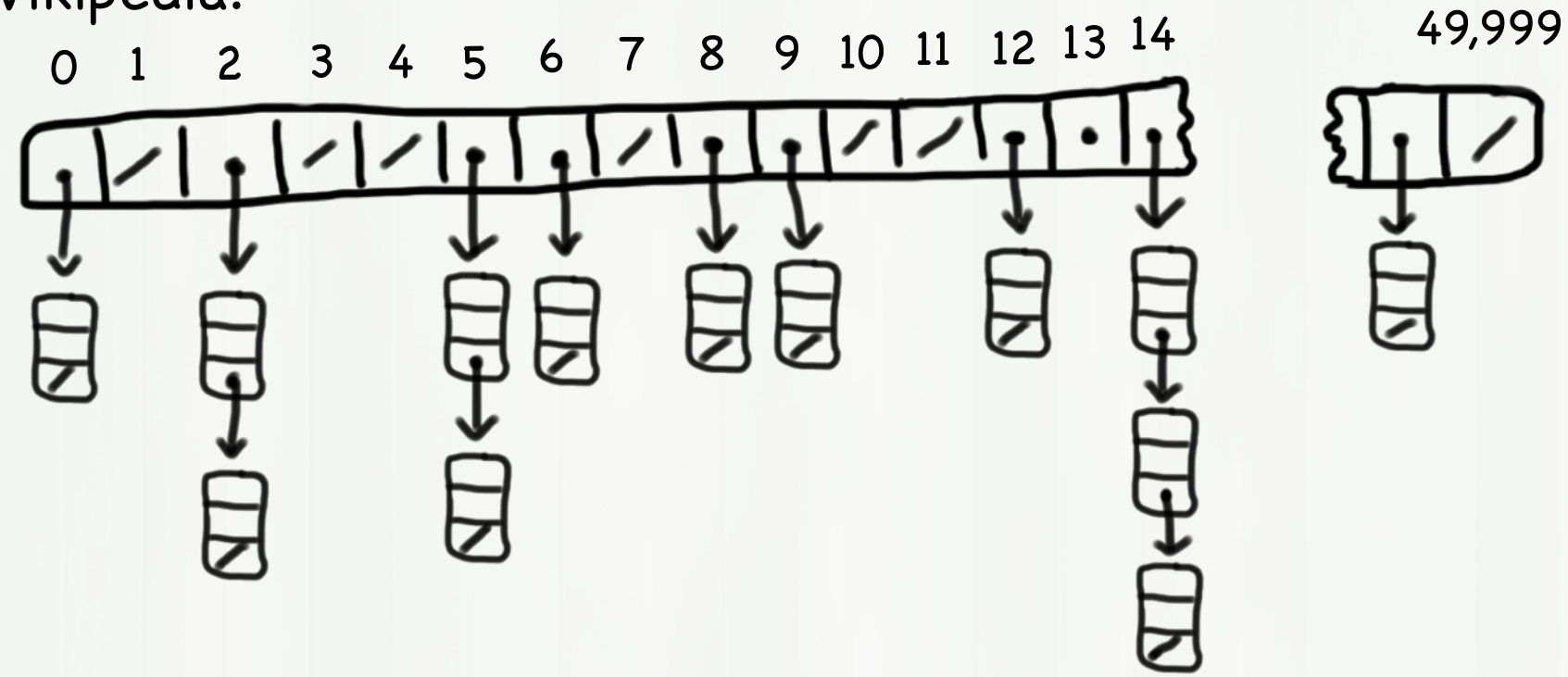
Wikipedia:



Hash Fn

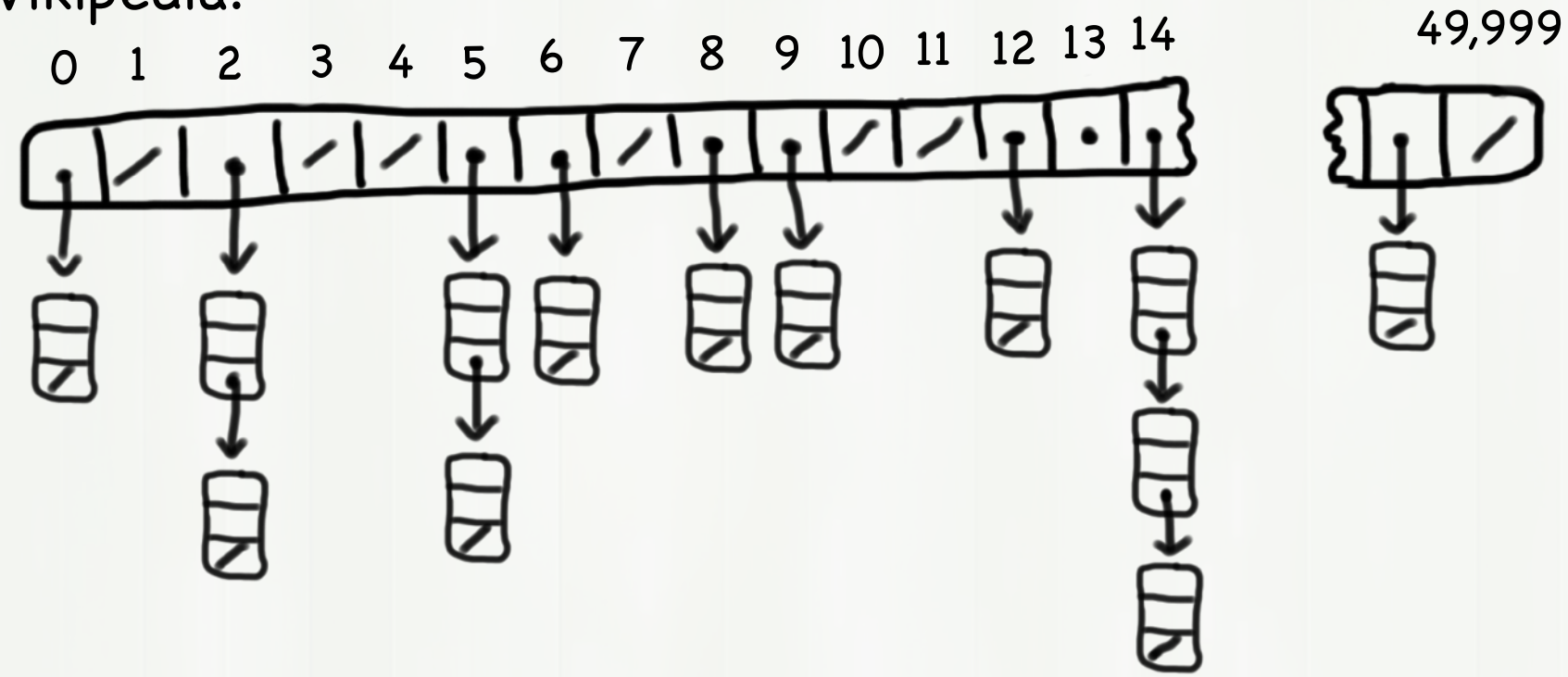
put("John Coltrane", html)

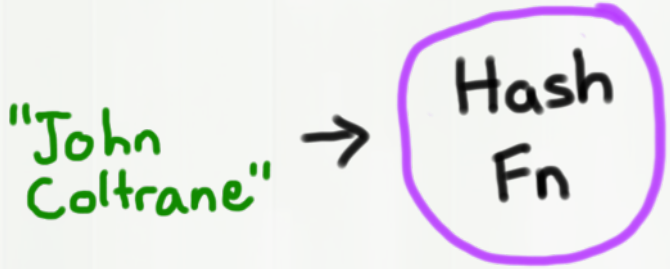
Wikipedia:



put("John Coltrane", html)

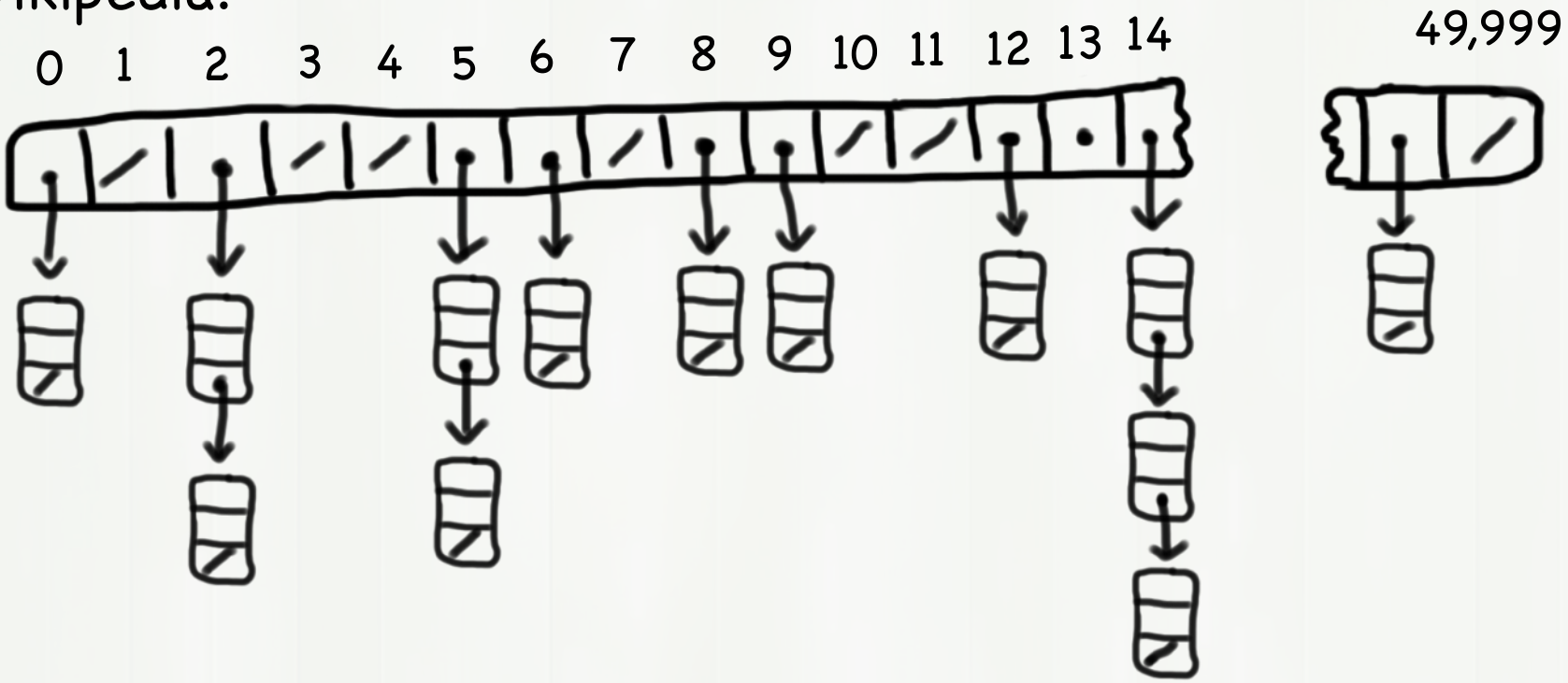
Wikipedia:





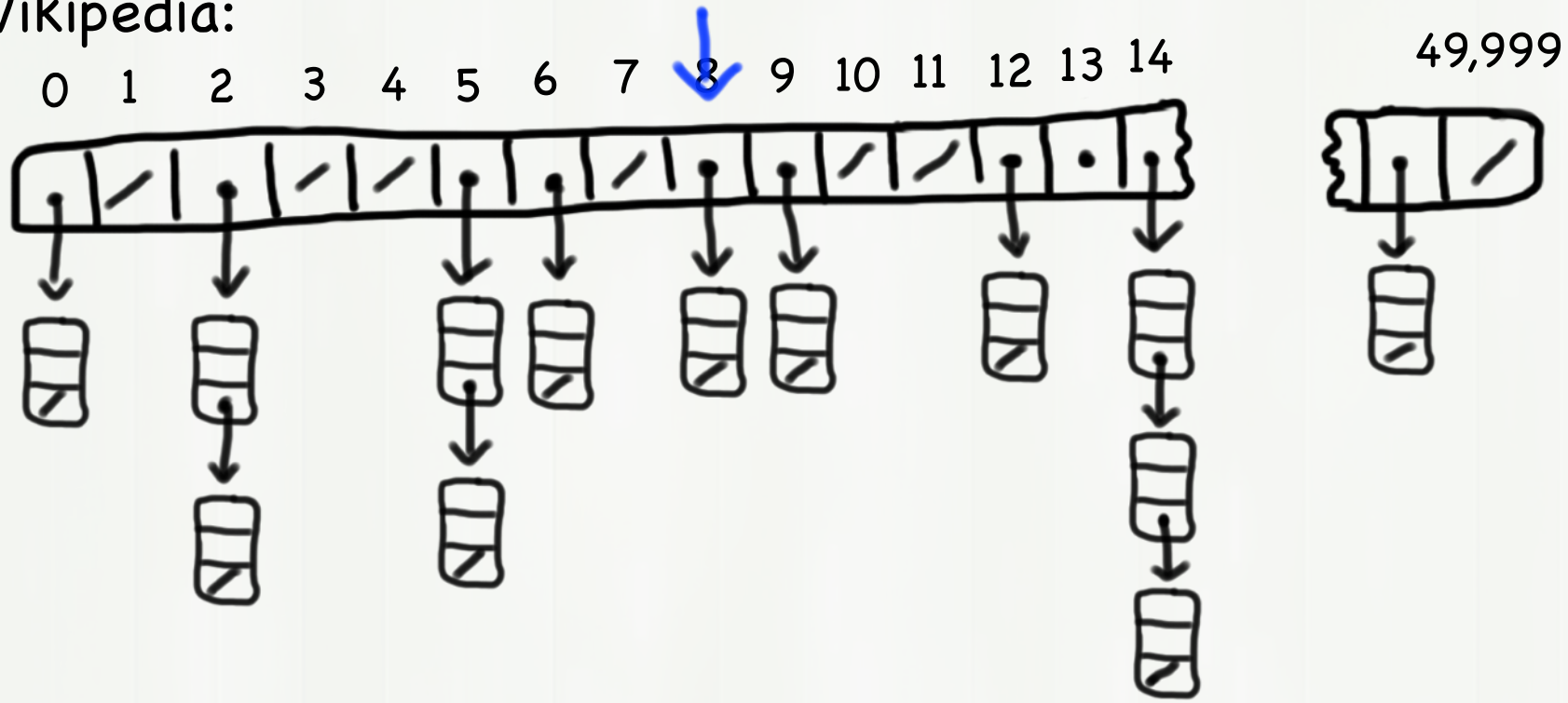
put("John Coltrane", html)

Wikipedia:



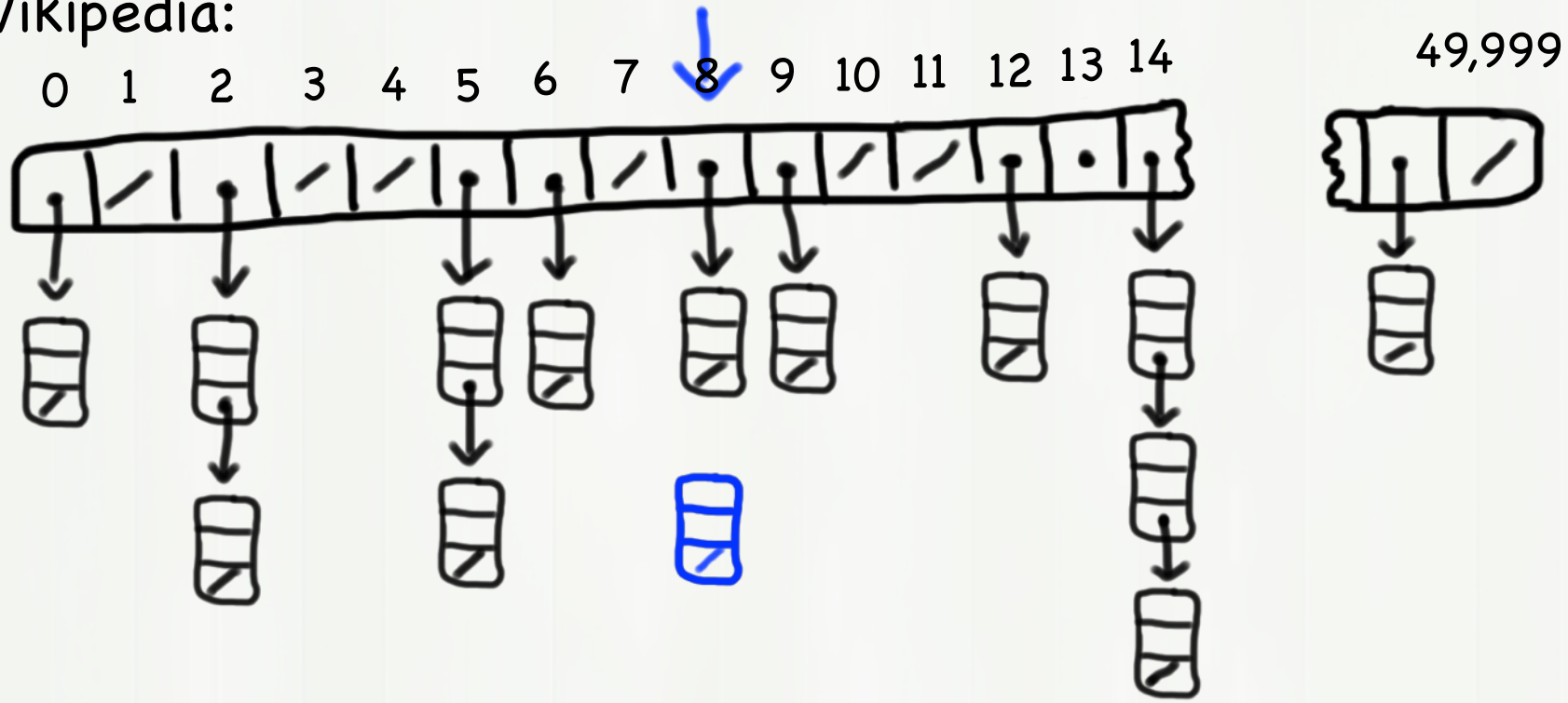
put("John Coltrane", html)

Wikipedia:



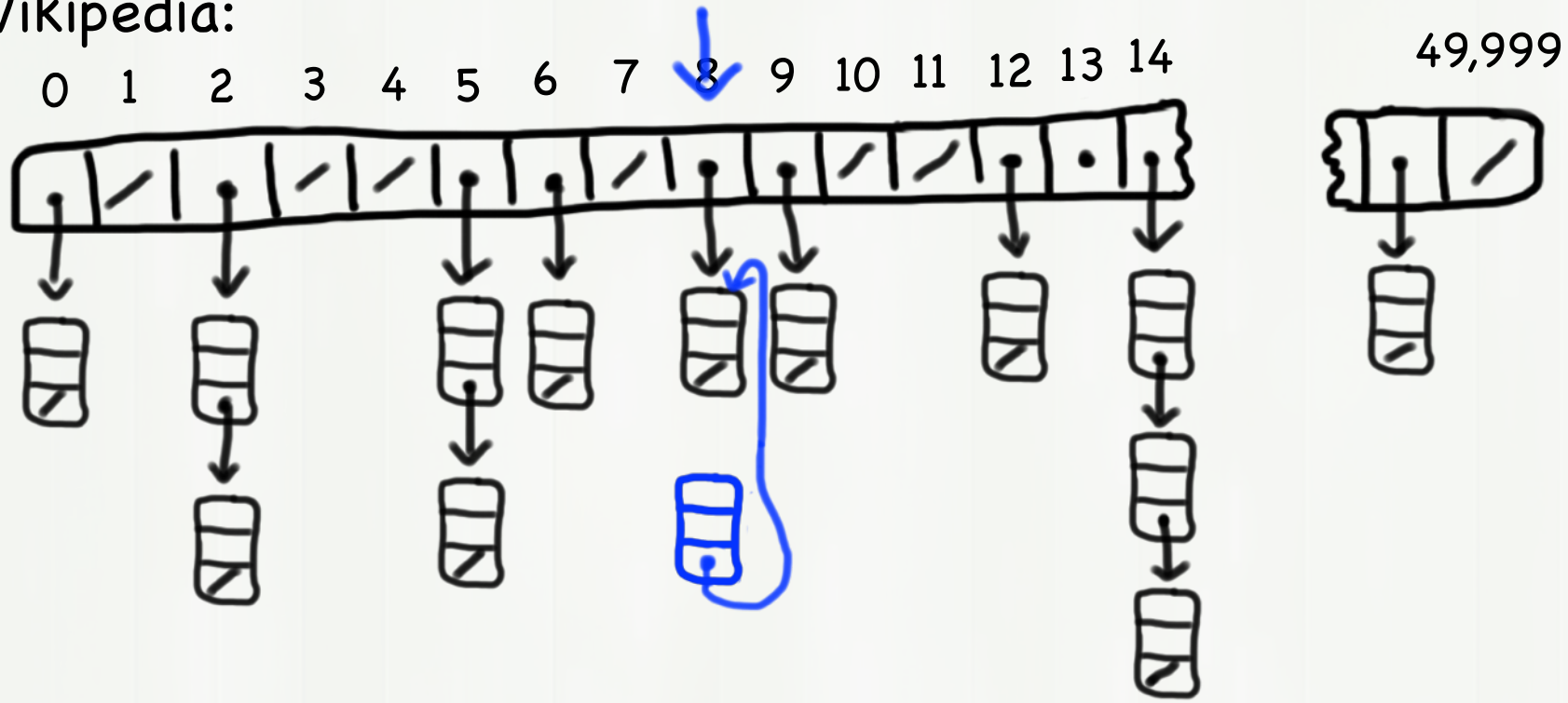
put("John Coltrane", html)

Wikipedia:



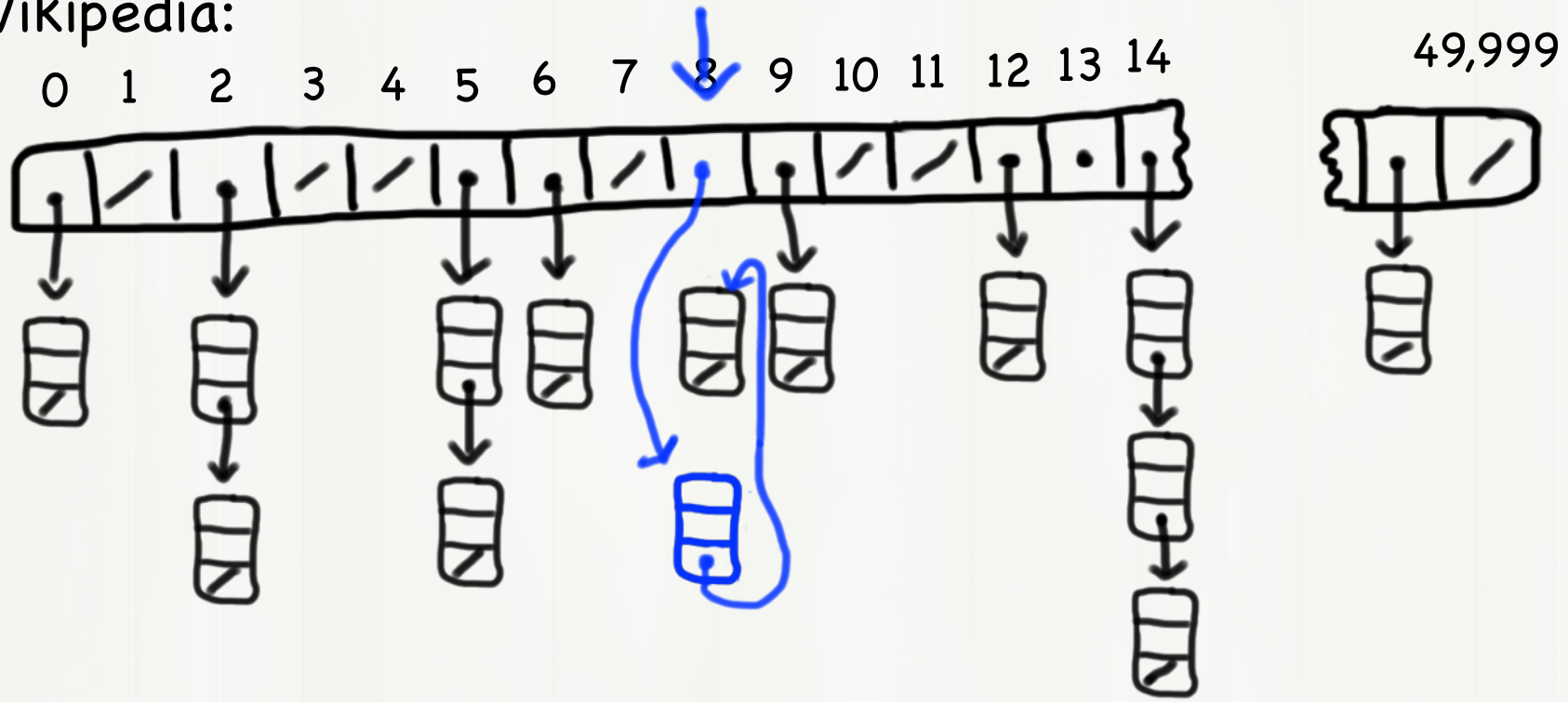
put("John Coltrane", html)

Wikipedia:



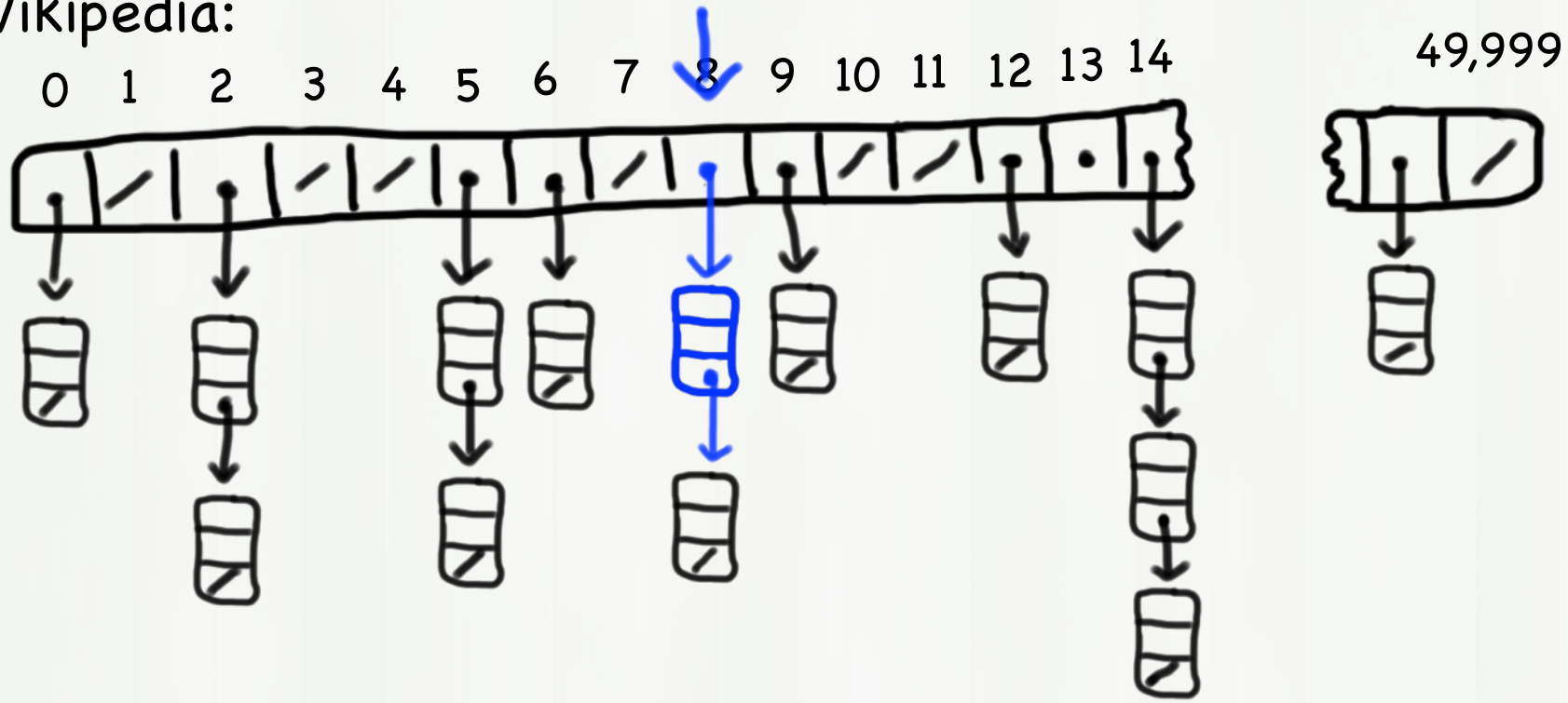
put("John Coltrane", html)

Wikipedia:



put("John Coltrane", html)

Wikipedia:



HashMap Algorithm Summary

Initialization:

0. Make a large array of linked lists

Getting and Putting:

1. Hash: key -> bucketIndex.
2. Jump directly to the bucket head.
3. Run linear search over the short list.

Today's Route

(eg HashMap<string, string>)

Other Applications

Code Wikipedia

Real Hash fns

Collisions

Hashing

HashMaps River

Today's Route

(eg HashMap<string, string>)

Other Applications

Code Wikipedia

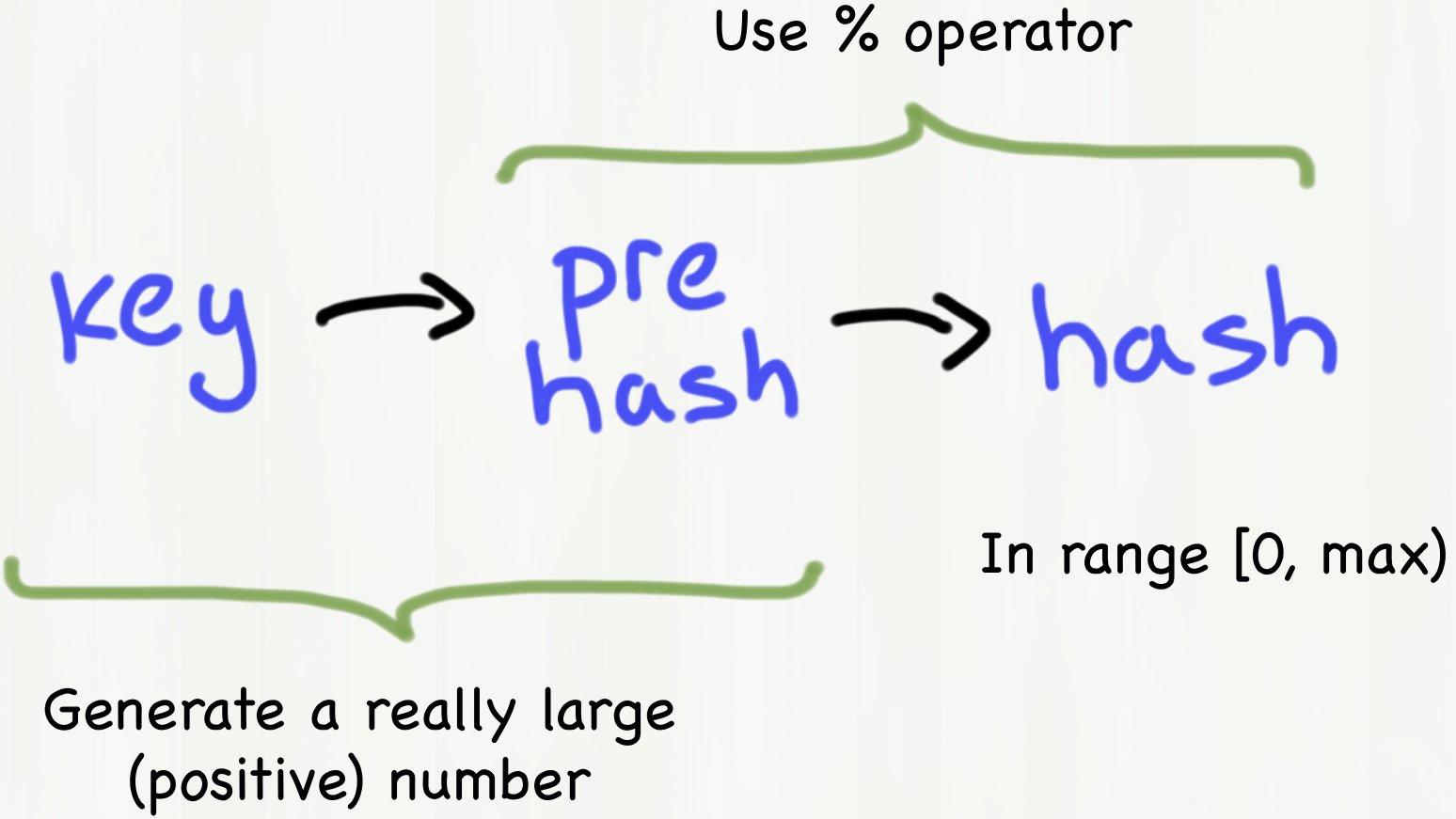
Real Hash fns

Collisions

Hashing

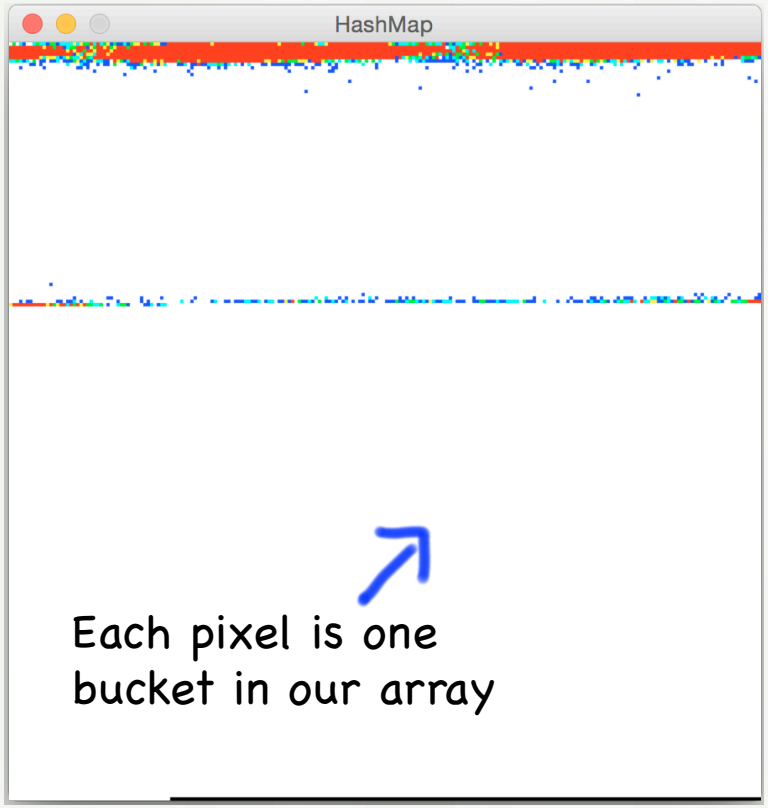
HashMaps River

Hash Function

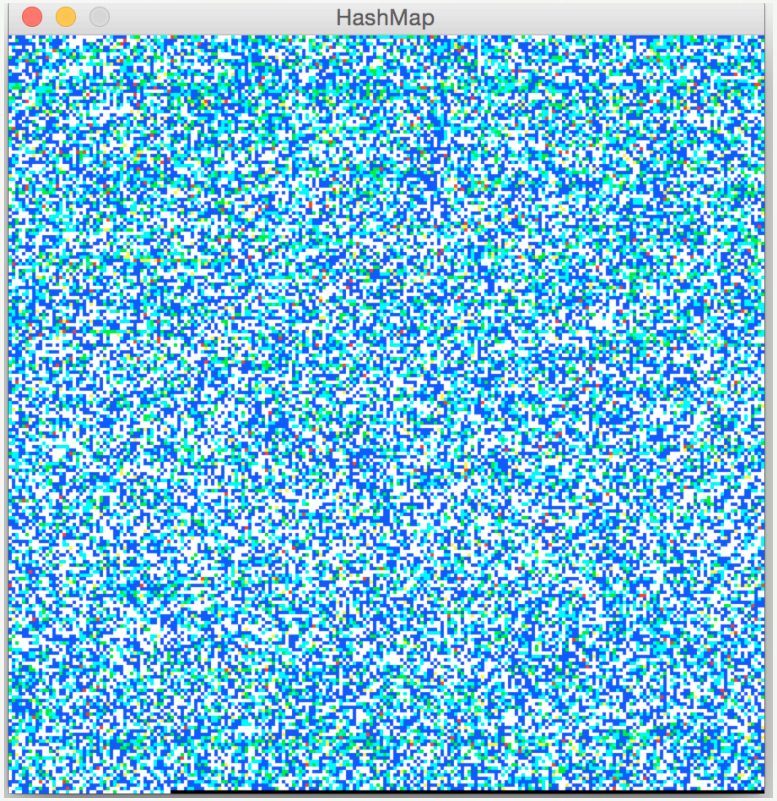


Judge Strings by the Contents of their Characters

preHash = Add each character in a string



preHash = Add each character in a string, weighted by 31^i



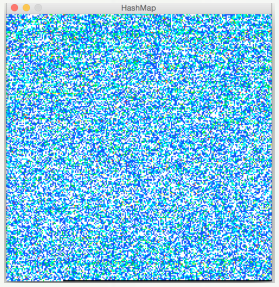
50,000 wikipedia articles

0 2 4

1 3 >= 5

Number of collisions

The Java Hash



Given string a as a key.

```
prehash =
```

$$a[0] + 31 \cdot a[1] + 31^2 \cdot a[2] + \cdots + 31^n \cdot a[n]$$

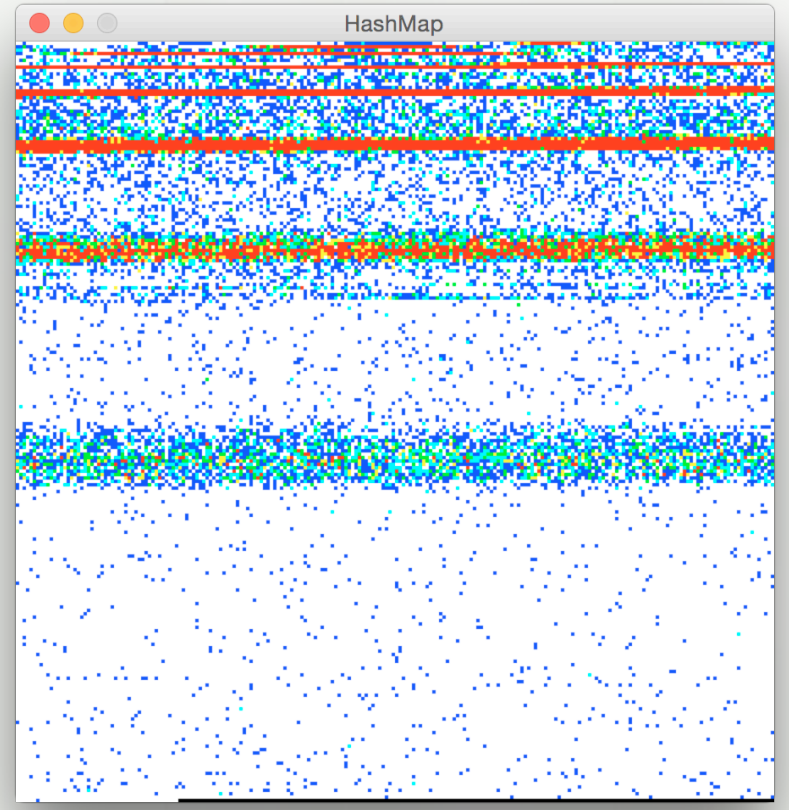
```
return (prehash % numBuckets)
```

* In java, the weights are in reverse order. That is not important :).

Why 31?

What if it was 2?

$$a[0] + 2 \cdot a[1] + 2^2 \cdot a[2] + \cdots + 2^n \cdot a[n]$$



Bucket distribution of 50,000 article titles from Wikipedia

Question:

What is the big-O of the **`get`** method?

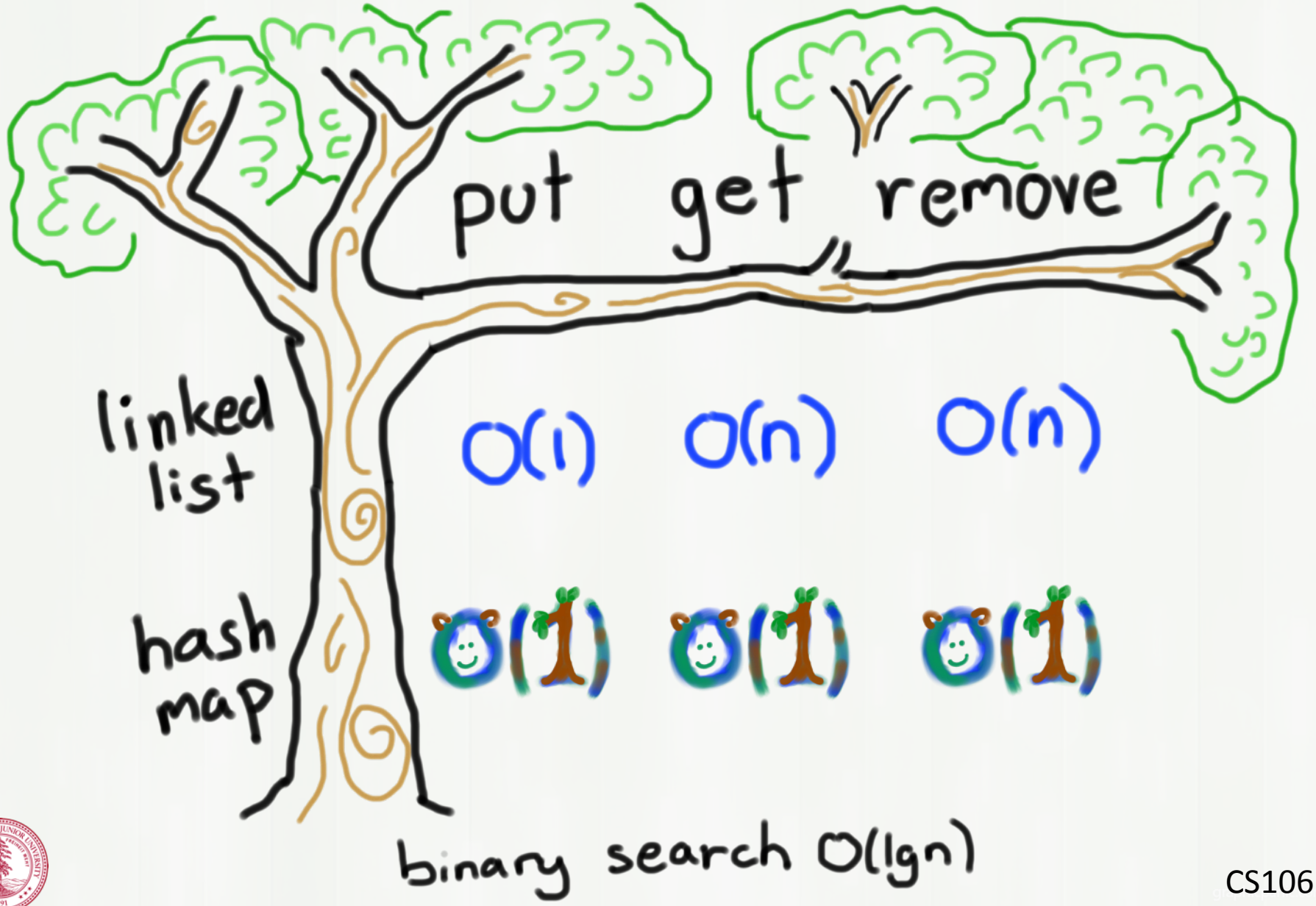
[suspense]

Answer:

* In expectation

What if nBuckets is too small?!?

HashMap Big-O



Knead to Program It

Gopher It

Source: Planet Animal Zone

Today's Route

(eg HashMap<string, string>)

Other Applications

Code Wikipedia

Real Hash fns

Collisions

Hashing

HashMaps River

Today's Route

(eg HashMap<string, string>)

Other Applications

Code Wikipedia

Real Hash fns

Collisions

Hashing

HashMaps River

Today's Route

(eg HashMap<string, string>)

Other Applications

Code Wikipedia

Real Hash fns

Collisions

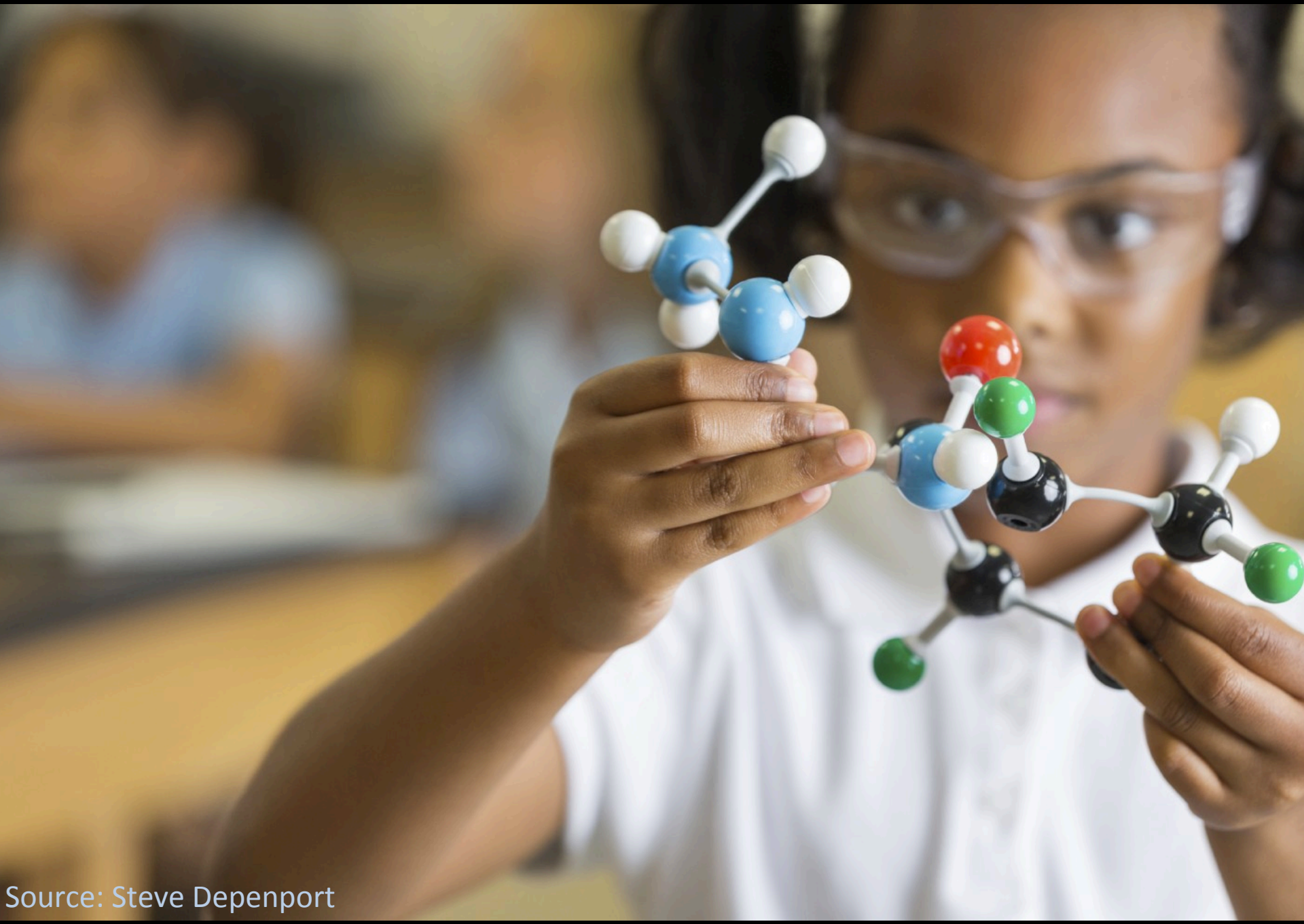
Hashing

HashMaps River

And Here We Are

Shazam

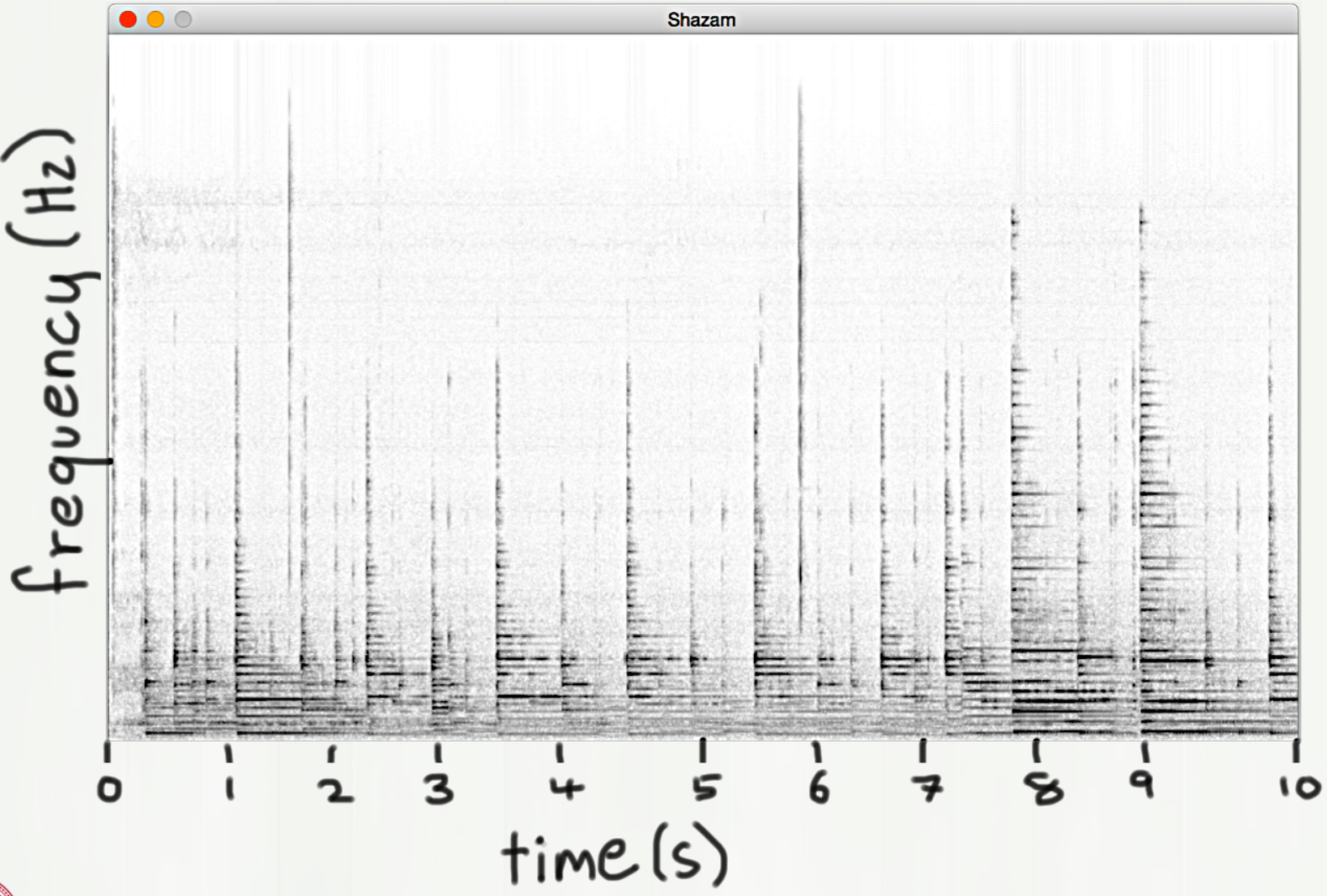
Source: Shazam



Source: Steve Depenport

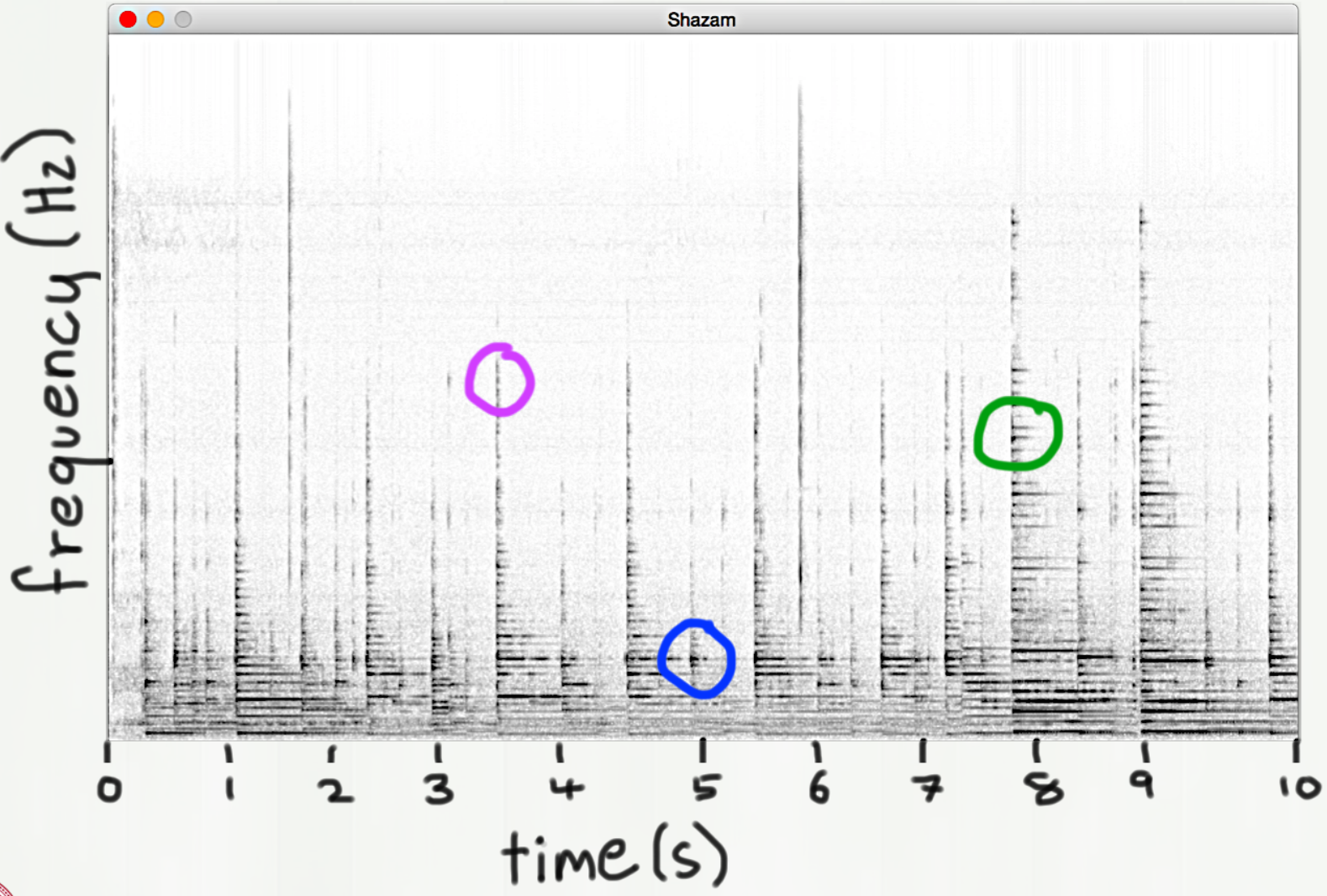
Similar to Wikipedia:
Huge search problem.

Spectrogram



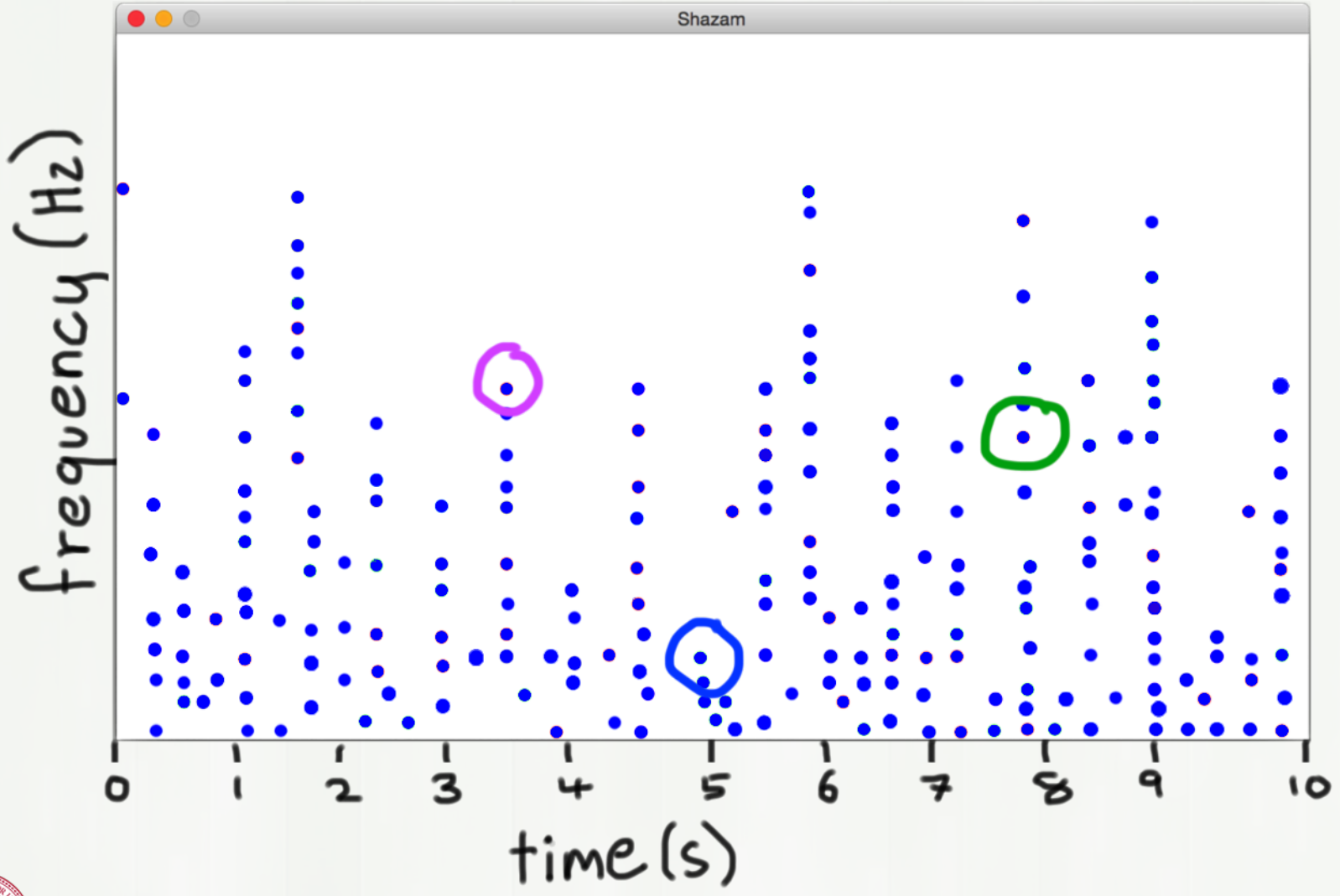
Does anyone recognize this song?

Spectrogram



Does anyone recognize this song?

Notes Over Time

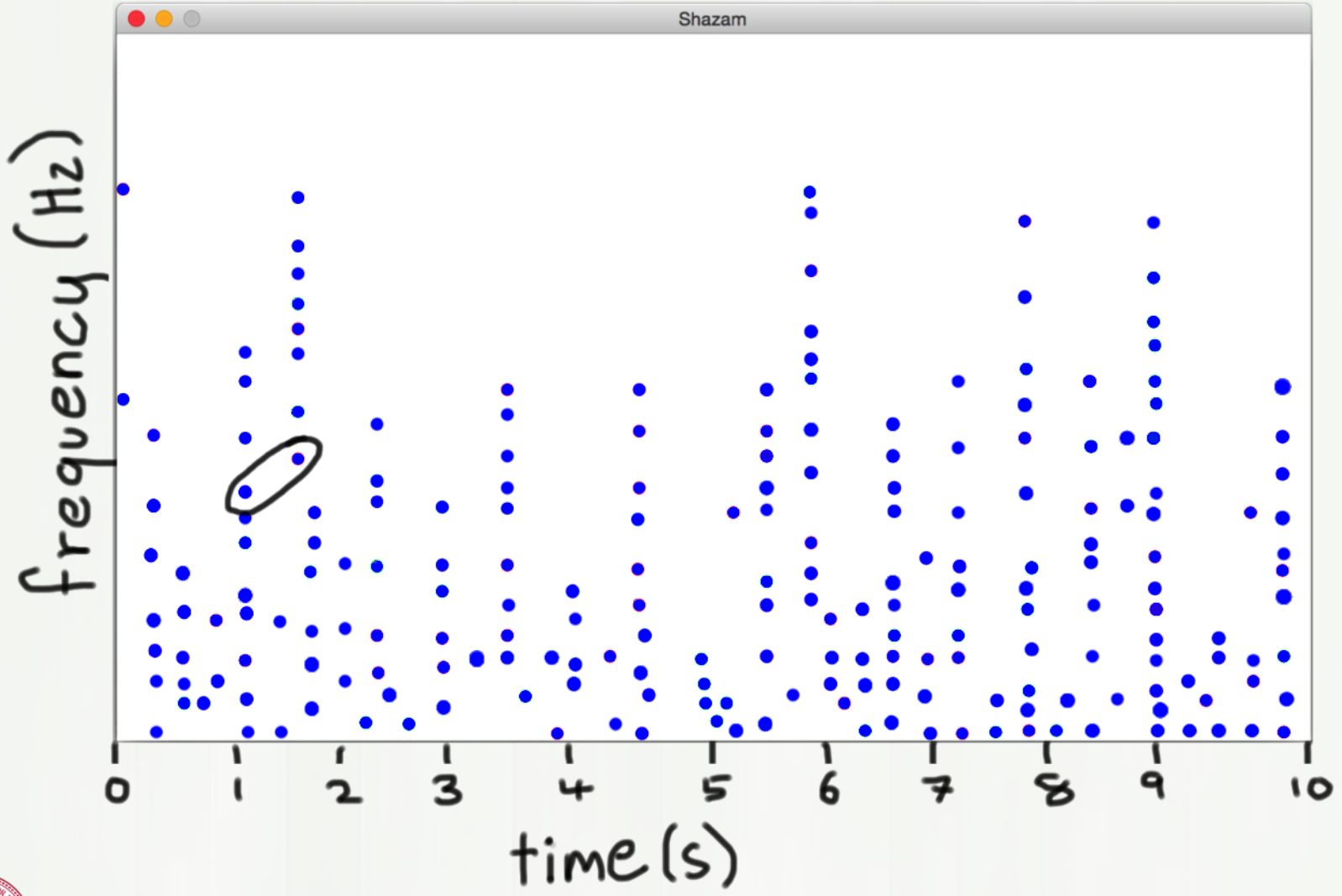


Wang, A. An Industrial-Strength Audio Search Algorithm

Hash the whole thing?

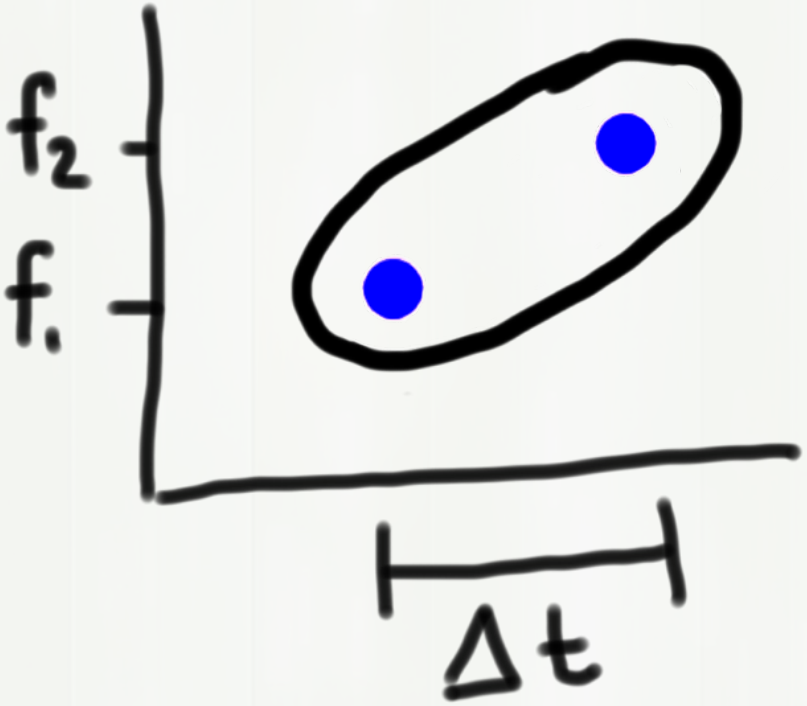
No

Find Pairs of Notes



Wang, A. An Industrial-Strength Audio Search Algorithm

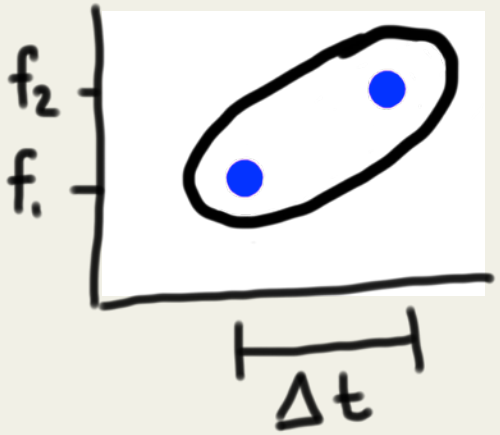
Note Pairs



Note/Songs Map

Shazam

Key:



Value:

You Can Call Me Al – Paul Simon. 7s,
You Can Call Me Al – Paul Simon. 43s,
All Right Now – Police. 18s

Wikipedia

Key:

"Antelope Canyon"

Value:

Antelope Canyon is a slot canyon in the American Southwest. It is located on Navajo land east of Page, Arizona. Antelope Canyon includes two separate, photogenic slot canyon sections, referred to individually as *Upper Antelope Canyon* or *The Crack*; and *Antelope Canyon* or *The Corkscrew*.[2]

The Navajo name for Upper Antelope Canyon is Tsé bighánílíní, which means "the place where water runs through rocks." Lower Antelope Canyon is Hazdistazí (advertised as *"Hasdestwazi"* by the Navajo Parks and Recreation Department), or "spiral rock arches." Both are located within the LeChee Chapter of the Navajo Nation.[4]

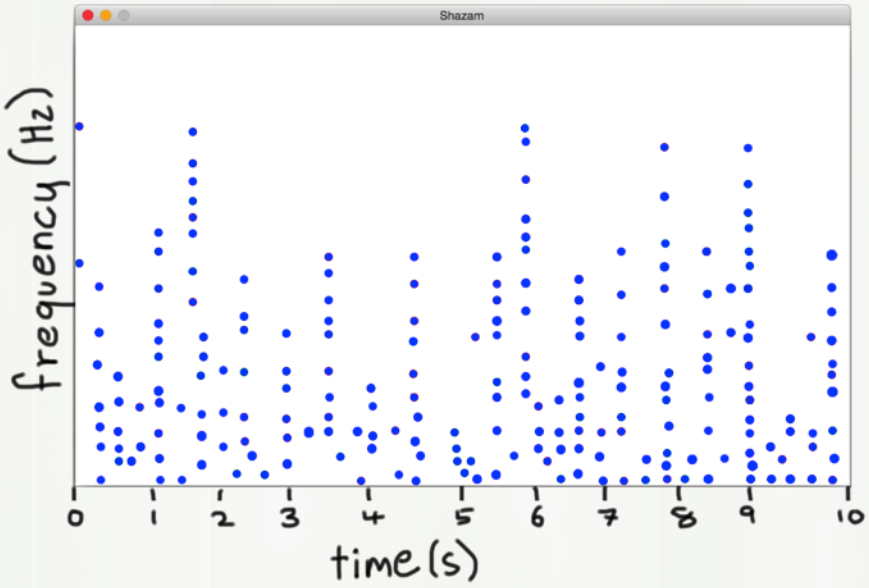
Contents [hide]
1 Geology
2 Tourism and photography
2.1 Upper Antelope Canyon

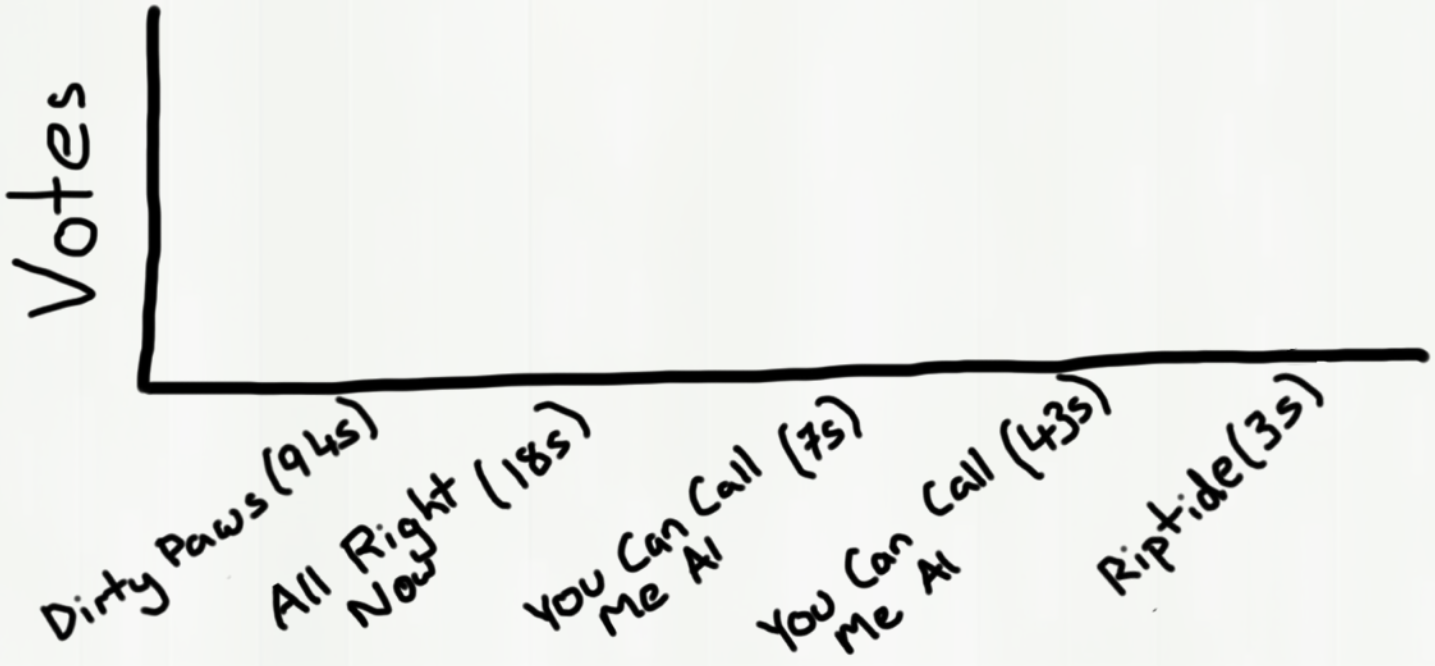
Antelope Canyon
Tsé bighánílíní dóó Hazdistazí (Navajo)

A beam of light in Upper Antelope Canyon

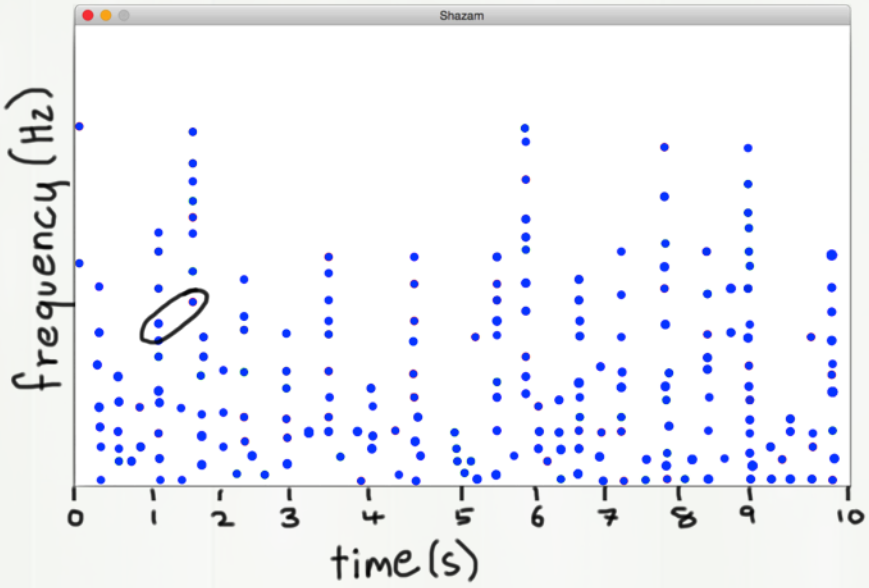
Antelope Canyon

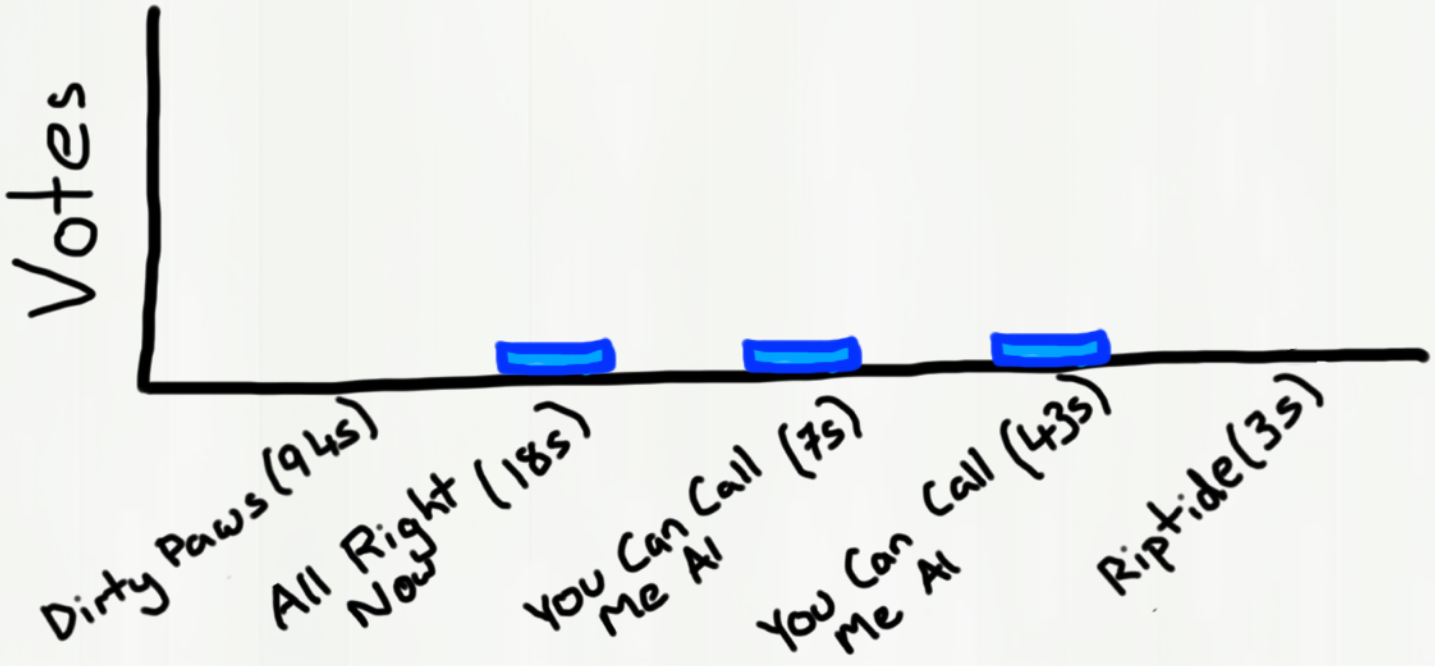
Note Pair Hashing



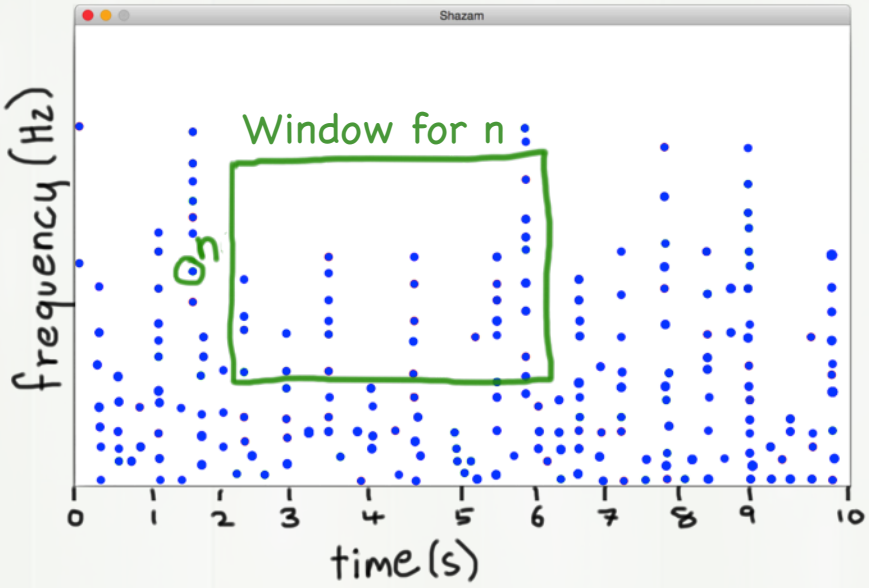


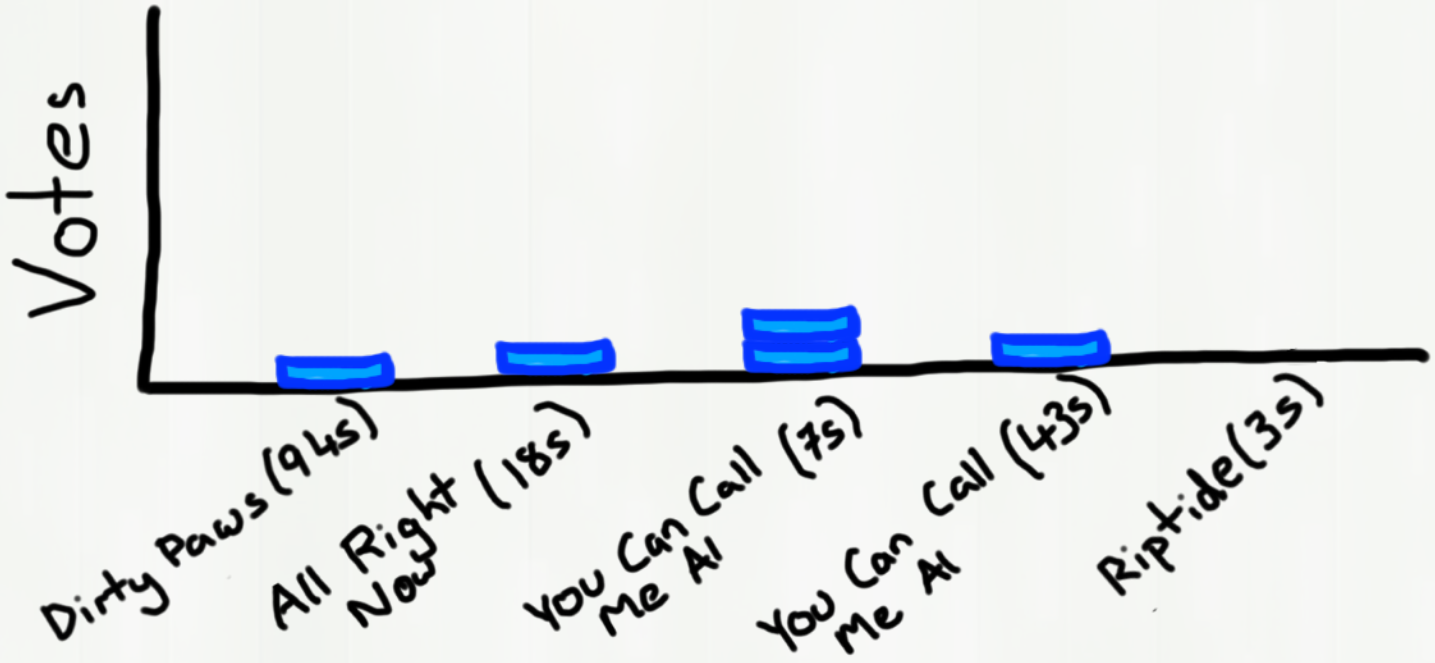
Note Pair Hashing



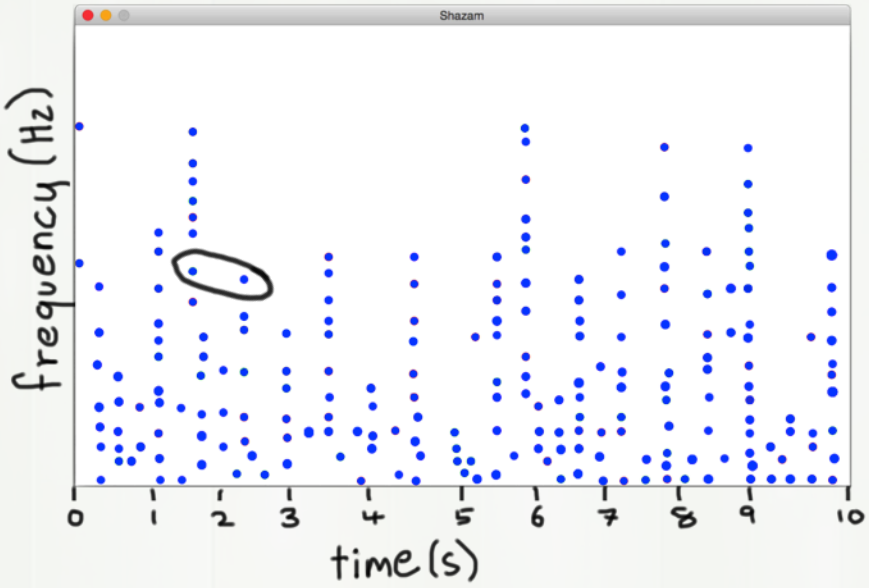


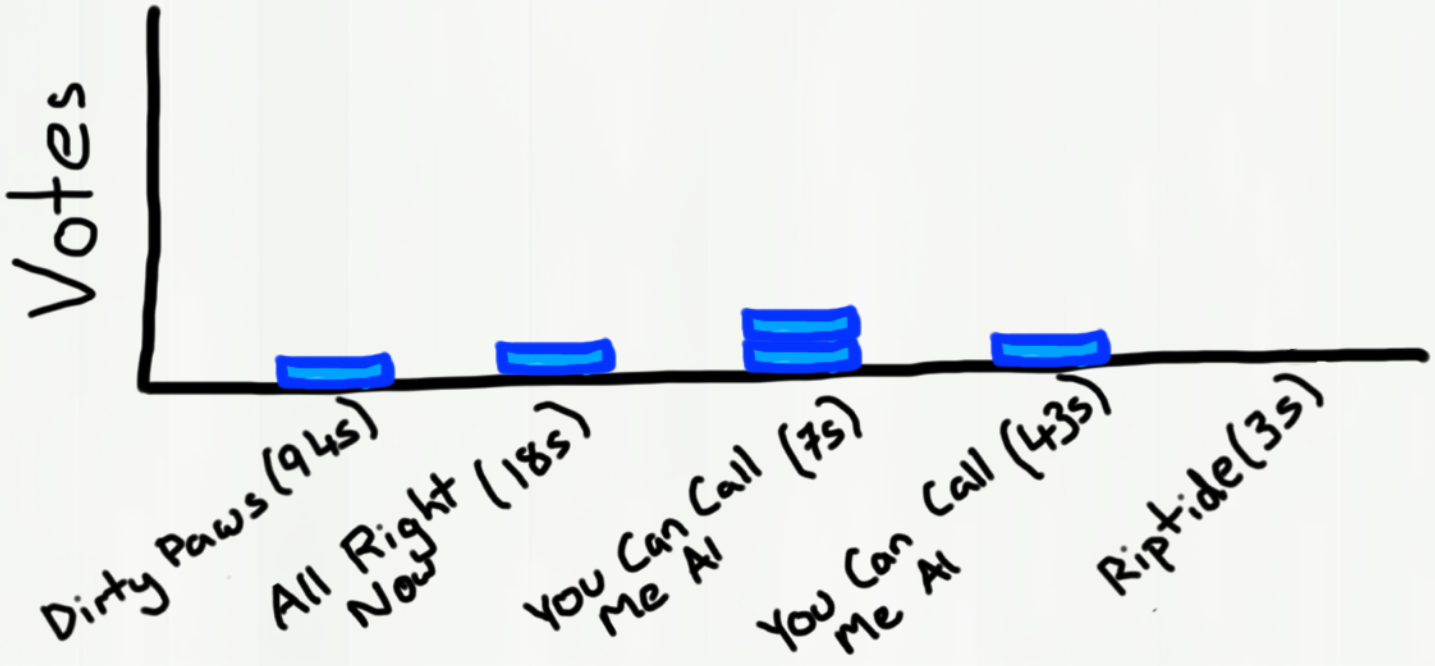
Note Pair Hashing



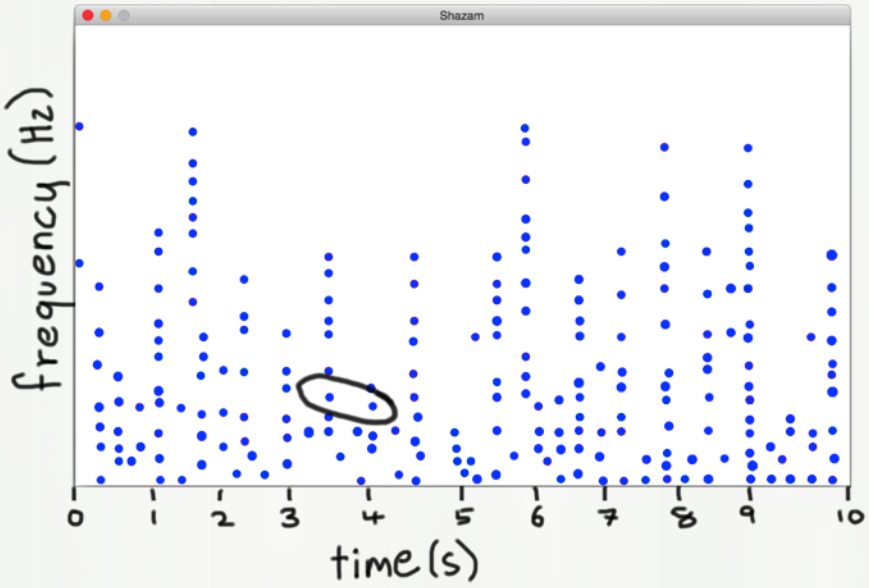


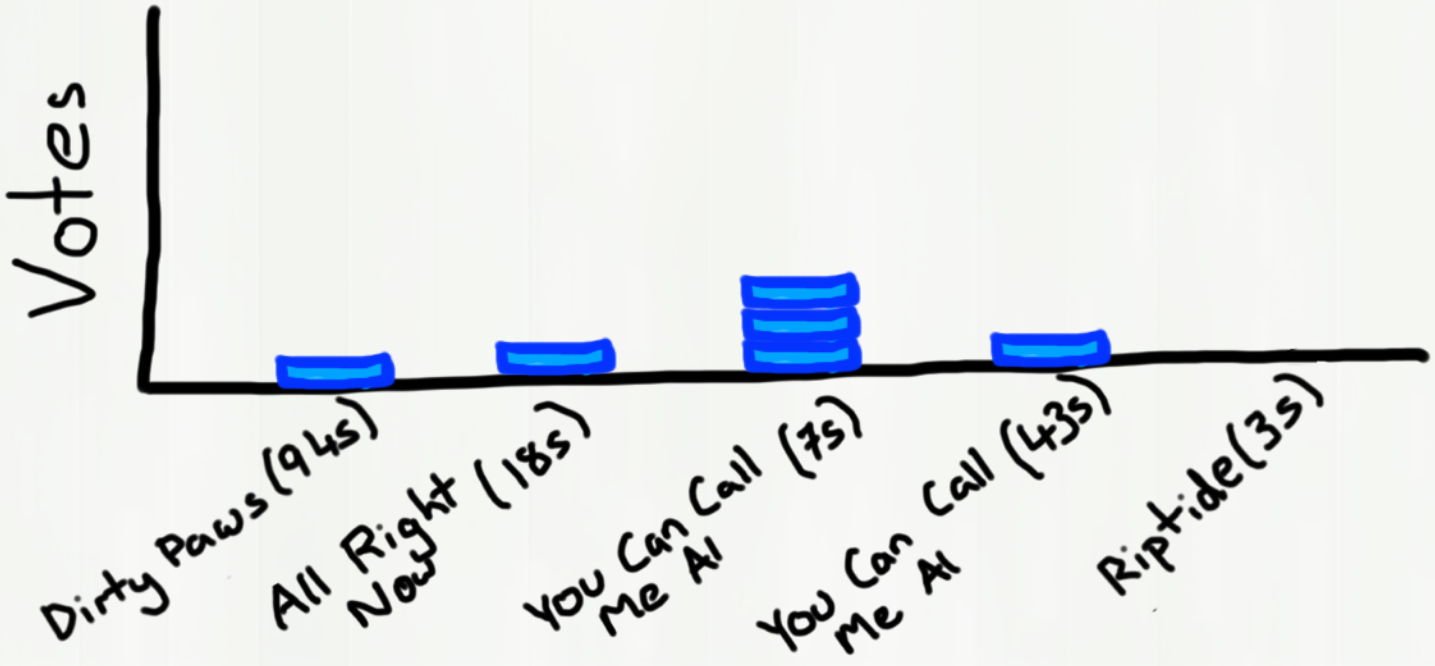
Note Pair Hashing



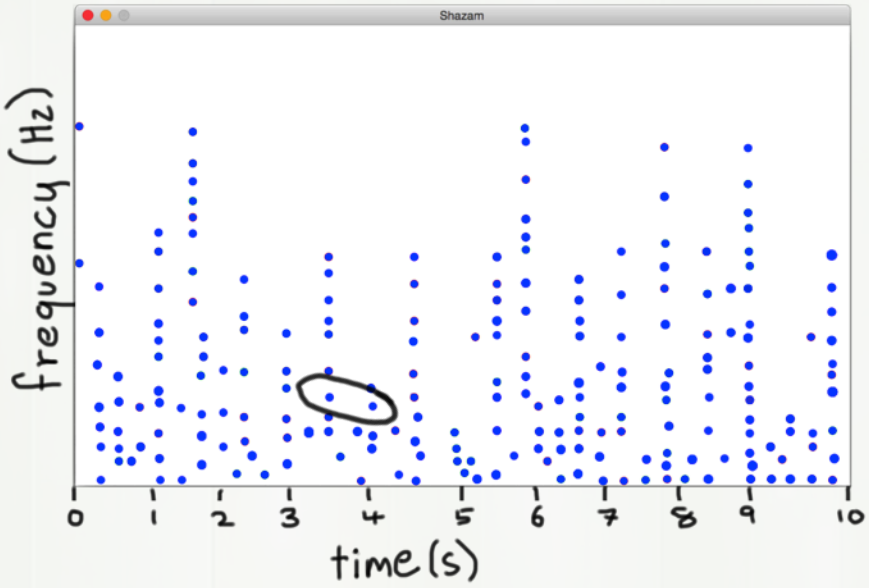


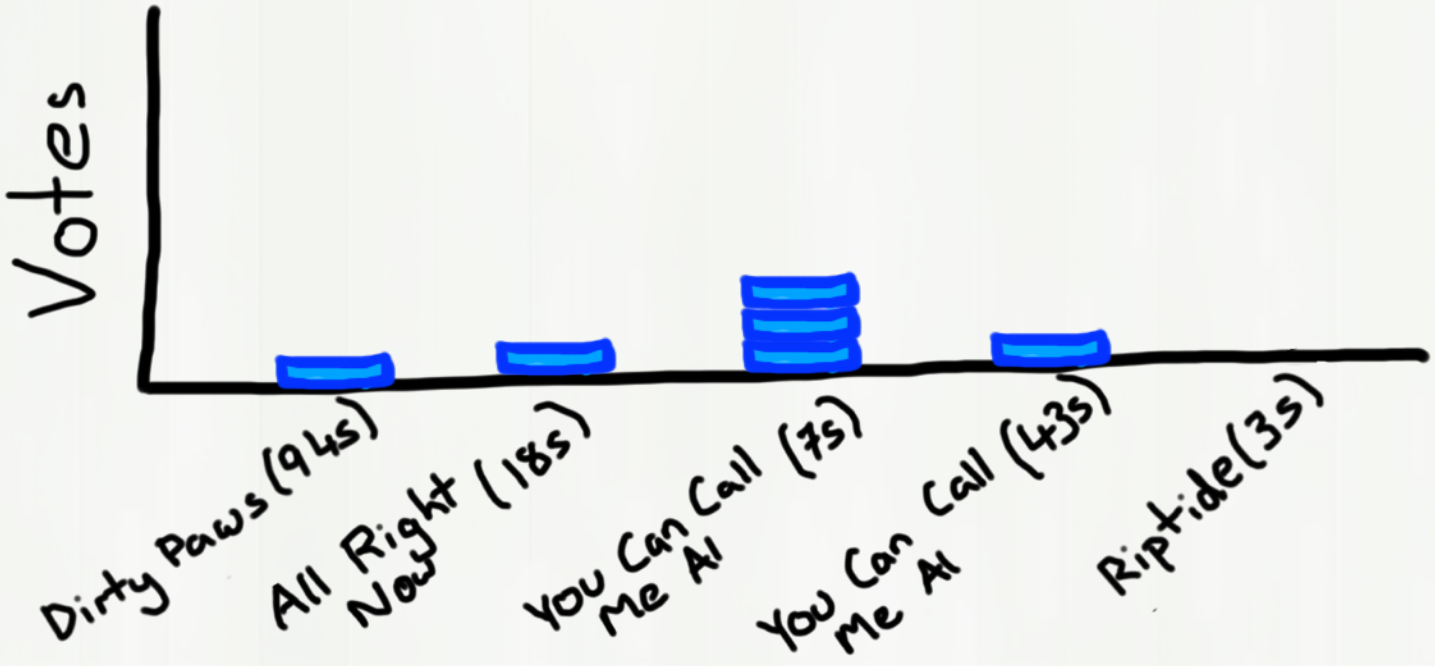
Note Pair Hashing



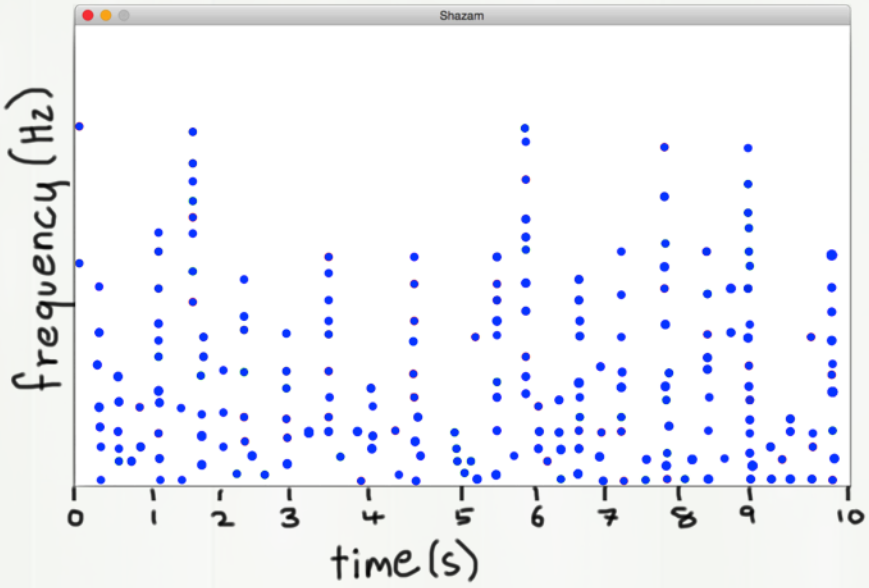


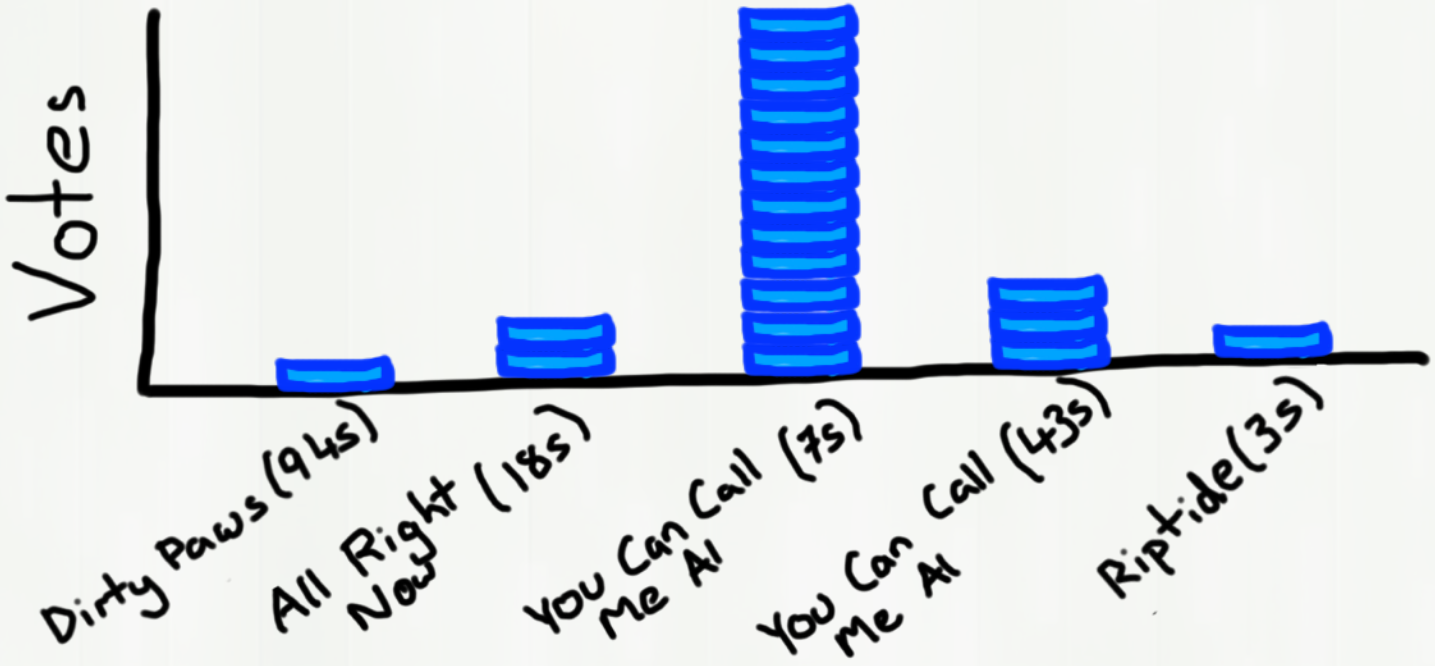
Note Pair Hashing



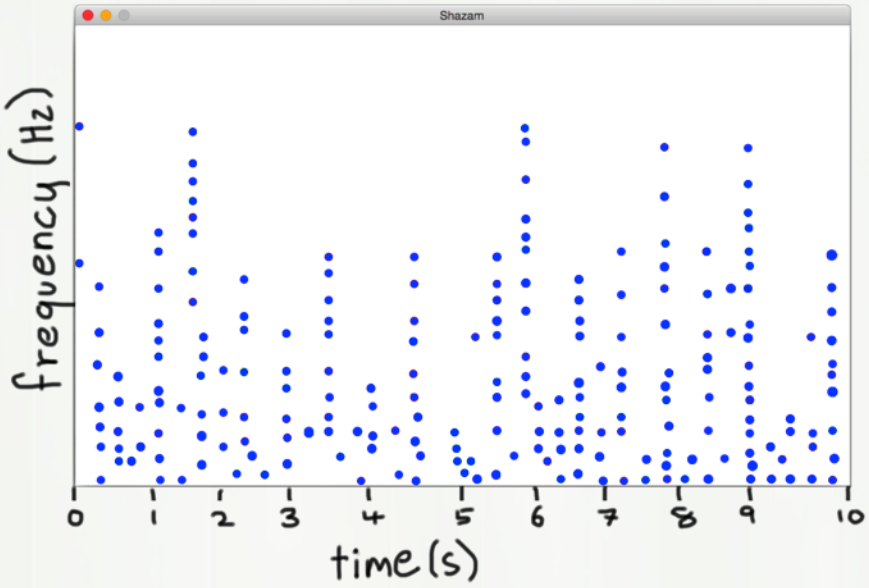


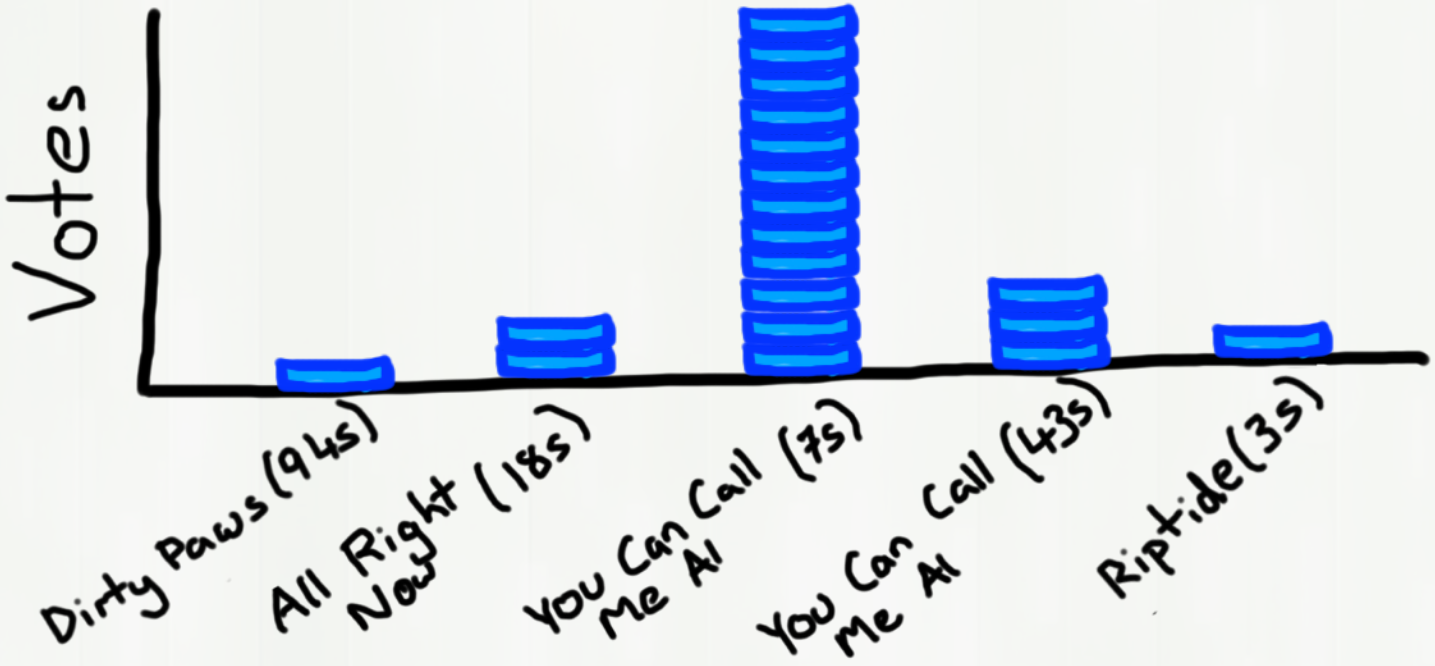
Note Pair Hashing



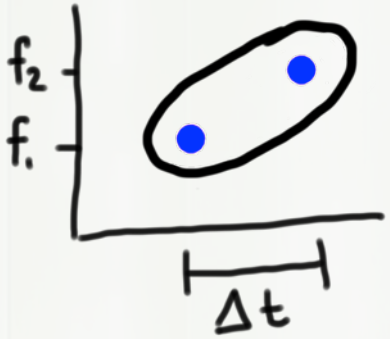


Note Pair Hashing





Hashing Note Pairs




```
int hash(int f1, int f2, int timeDelta) {
    int p = 31;
    int pre = f1 + (p * f2) + (p * p * timeDelta);
    return pre % NUM_BUCKETS;
}
```

You Can Call Me Al – Paul Simon. 23s,
You Can Call Me Al – Paul Simon. 54s,
Message in a Bottle – Police. 92s

Code Available

Review

HashMap Algorithm Summary

Initialization:

0. Make a large array of linked lists

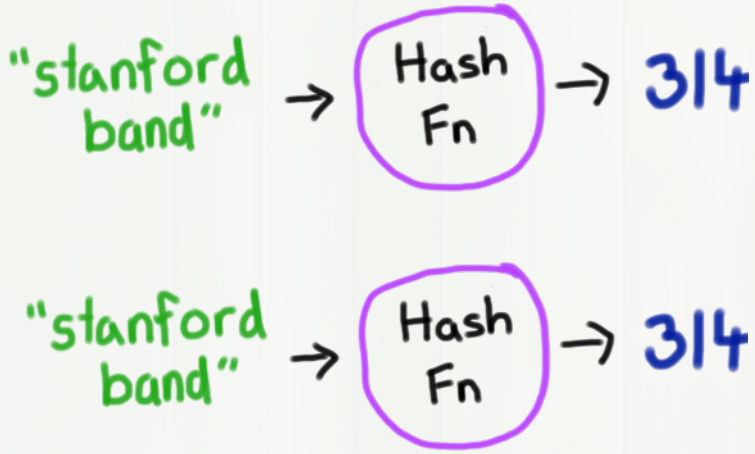
Getting and Putting:

1. Hash: key -> bucketIndex.
2. Jump directly to the bucket head.
3. Run simple search over the list.

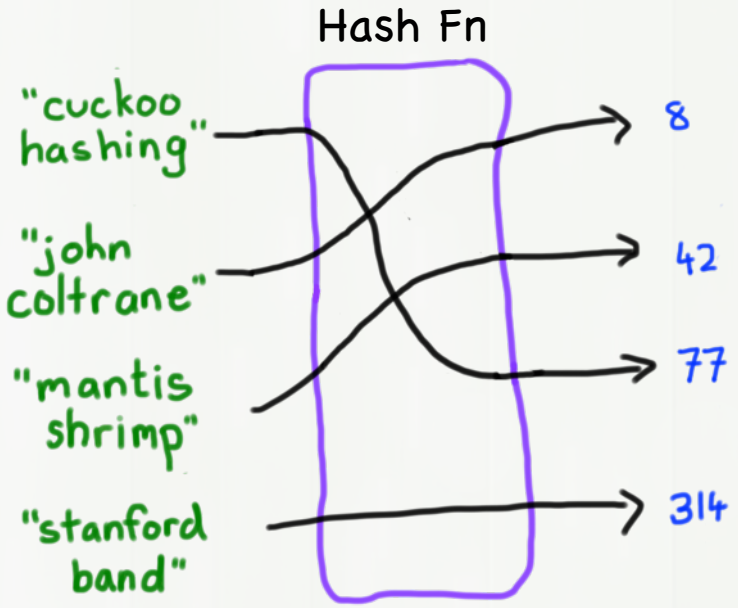
Hash Function

```
int hash(string key);
```

1. Consistent



2. Well Distributed



Today's Goals

1. Understand how the HashMap works
2. Add hashing to your algorithmic toolbox
3. Practice linked lists

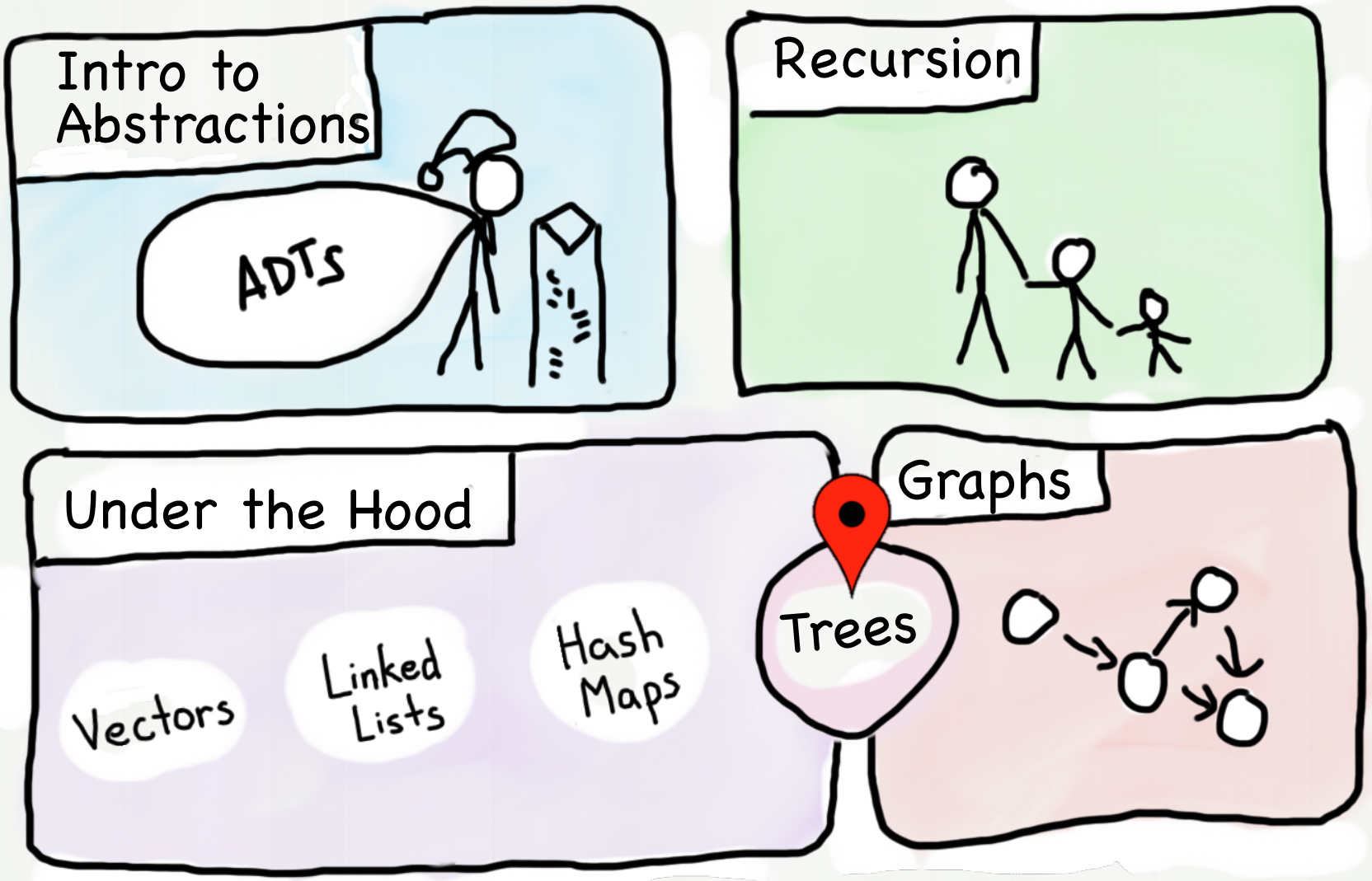
Today's Goals

1. Understand how the HashMap works
2. Add hashing to your algorithmic toolbox
3. Practice linked lists

Place and Time

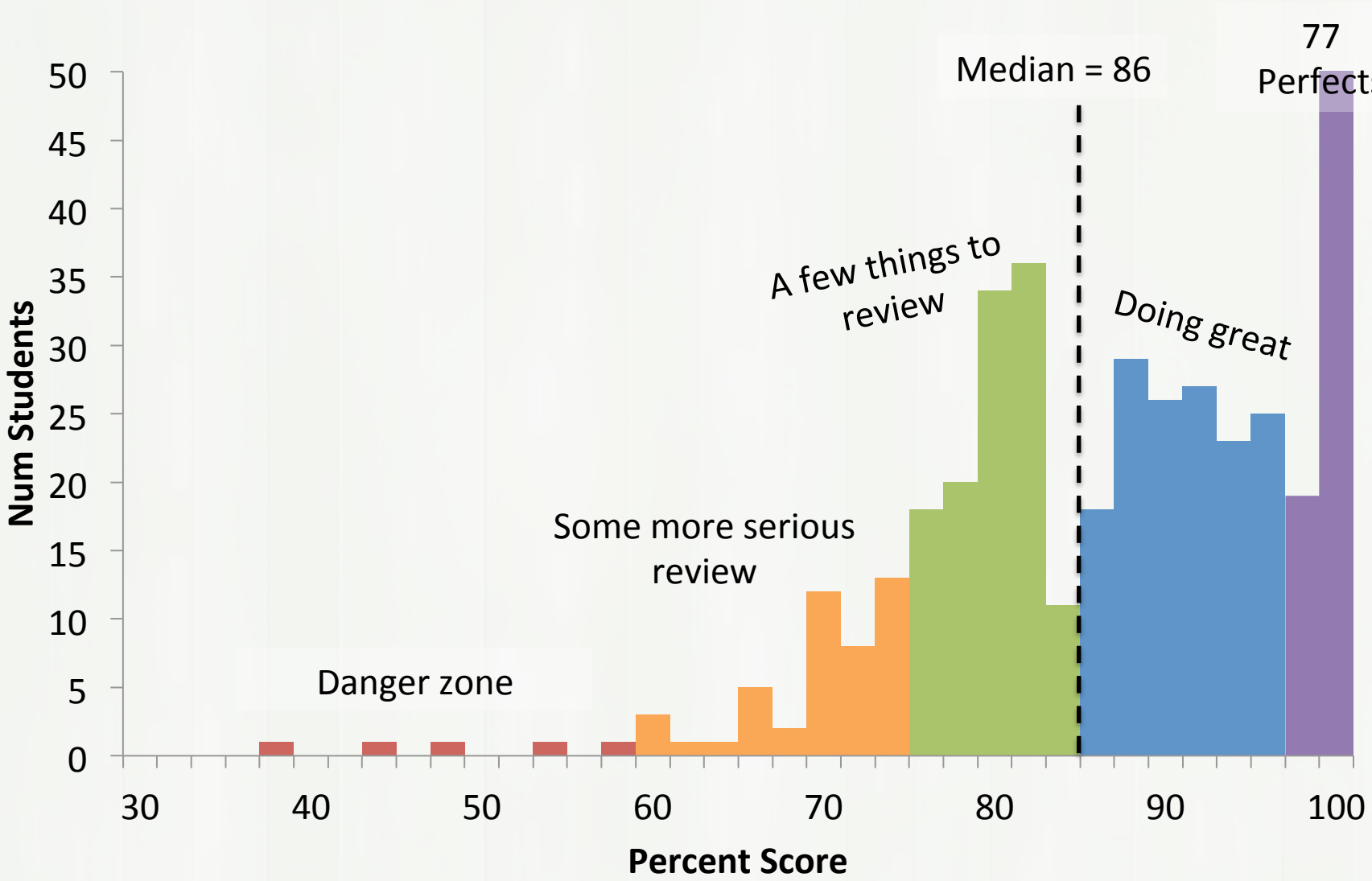
N Andy, P Chris, J Huang, L Guibas. Scalable Homework Search for Massive Open Online Programming Courses. WWW, Seoul 2014.

Friday Trees

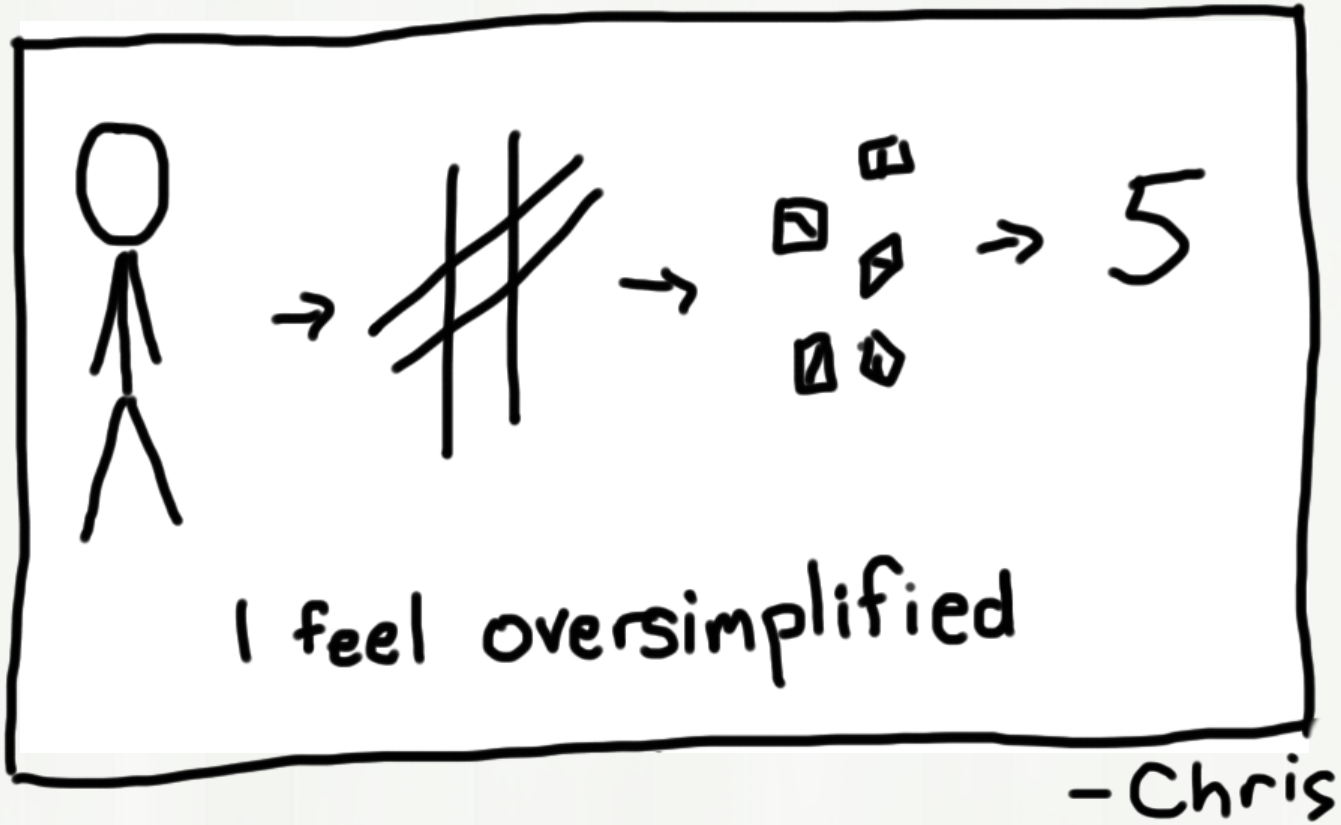


Thank you

Midterm


Median = 86
77
Perfects
A few things to review
Doing great
Some more serious review
Danger zone
Num Students
0
5
10
15
20
25
30
35
40
45
50
30
40
50
60
70
80
90
100
Percent Score

Modern Times



Details

Midterm
Review Policy
Answer Key