[301] Dictionary Nesting

Tyler Caraza-Harter

Learning Objectives Today

More dictionary operations

- len, in, for loop
- d.keys(), d.values()
- defaults for get and pop

Syntax for nesting (dicts inside dicts, etc)

- indexing/lookup
- step-by-step resolution

dict dict dict

Understand common use cases for nesting

- binning/bucketing (list in dict)
- a more convenient table representation (dict in list)
- transition probabilities with Markov chains (dict in dict)

one of the most common data analysis tasks

we'll generate random English-like texts

Today's Outline

Dictionary Ops

Binning (dict of list)

Table Representation (list of dict)

Probability Tables and Markov Chains (dict of dict)

Creation of Empty Dict

```
Non-empty dict:
d = {"a": "alpha", "b": "beta"}
Empty dict (way I):
d = \{\}
Empty dict (way 2):
d = dict()
  similar for lists: L = list()
   similar for sets: s = set()
```

```
num words = {0:"zero", 1:"one", 2:"two", 3:"three"}
print(len(num_words))
print(1 in num words)
print("one" in num words)
for x in num words:
    print(x)
```

```
num_words = {0:"zero", 1:"one", 2:"two", 3:"three"}
print(len(num words))
print(1 in num words)
print("one" in num words)
for x in num words:
    print(x)
```

```
num words = {0:"zero", 1:"one", 2:"two", 3:"three"}
print(len(num words))
print(1 in num words)
                                    True
                                      False
print("one" in num words)
                                      (it is only checking keys, not vals)
for x in num words:
    print(x)
```

```
num words = {0:"zero", 1:"one", 2:"two", 3:"three"}
print(len(num words))
print(1 in num words)
                                    True
                                      False
print("one" in num words)
                                      (it is only checking keys, not vals)
for x in num words:
    print(x)
```

```
num_words = {0:"zero", 1:"one", 2:"two", 3:"three"}
print(len(num words))
print(1 in num words)
                                        True
                                          False
print("one" in num words)
                                          (it is only checking keys, not vals)
for x in num words:
    print(x)
                                         (for iterates over keys, not vals)
                                         (note there is no order here)
```

by combining a for loop with lookup

```
num words = {0:"zero", 1:"one", 2:"two", 3:"three"}
print(len(num words))
print(1 in num words)
                                      True
                                       False
print("one" in num words)
                                       (it is only checking keys, not vals)
for x in num words:
                                       2 two
    print(x, num words[x])
                                         one
                                       0 zero
                                       3 three
     you can iterate over values
```

Extracting keys and values

don't worry about these new types, because we can force them to be lists

Extracting keys and values

```
num words = {0:"zero", 1:"one", 2:"two", 3:"three"}
print(type(num words.keys()))
                                           <class 'dict_keys'>
print(type(num words.values()))
                                            <class 'dict_values'>
print(list(num words.keys()))
                                          [3, 1, 2, 0]
                                           ["one", "two",
print(list(num words.values()))
                                            "zero", "three"]
```

```
suffix = {1:"st", 2:"nd", 3:"rd"}
```

```
suffix.pop(0) # delete fails, because no key 0
```

suffix[4] # lookup fails because no key 4

specify a default if

key cannot be found

```
suffix = {1:"st", 2:"nd", 3:"rd"}
suffix.pop(0) # delete fails, because no key 0
suffix[4] # lookup fails because no key 4
suffix.get(4, "th") # returns "th" because no key 4
```

```
suffix = {1:"st", 2:"nd", 3:"rd"}
                    specify a default if
key cannot be found
suffix.pop(0) # delete fails, because no key 0
suffix[4] # lookup fails because no key 4
suffix.get(4, "th") # returns "th" because no key 4
               specify a default if
               key cannot be found
```

```
suffix = {1:"st", 2:"nd", 3:"rd"}
                         specify a default if
key cannot be found
suffix.pop(0, "th") # returns "th" because no key 0
suffix[4] # lookup fails because no key 4
suffix.get(4, "th") # returns "th" because no key 4
               specify a default if
              key cannot be found
```

```
suffix = {1:"st", 2:"nd", 3:"rd"}
for num in range(6):
    print(str(num) + suffix.get(num, "th"))
                    Oth
                    Ist
                    2nd
                    3rd
                    4th
                    5th
```

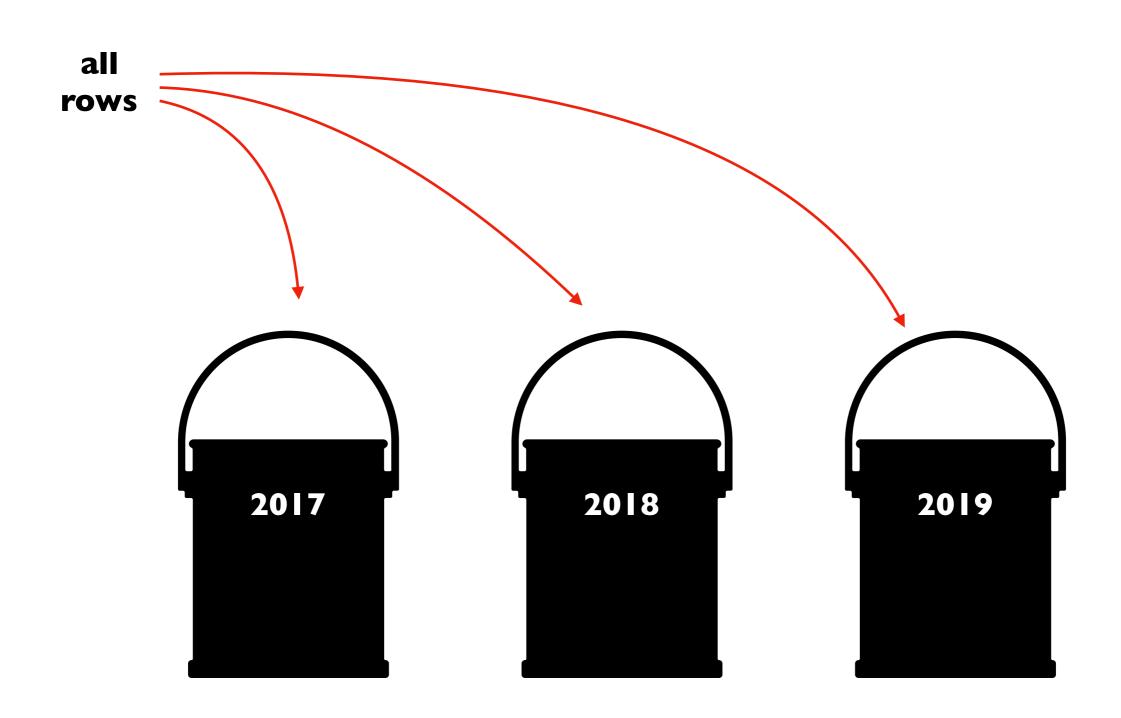
Today's Outline

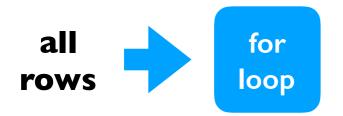
Dictionary Ops

Binning (dict of list)

Table Representation (list of dict)

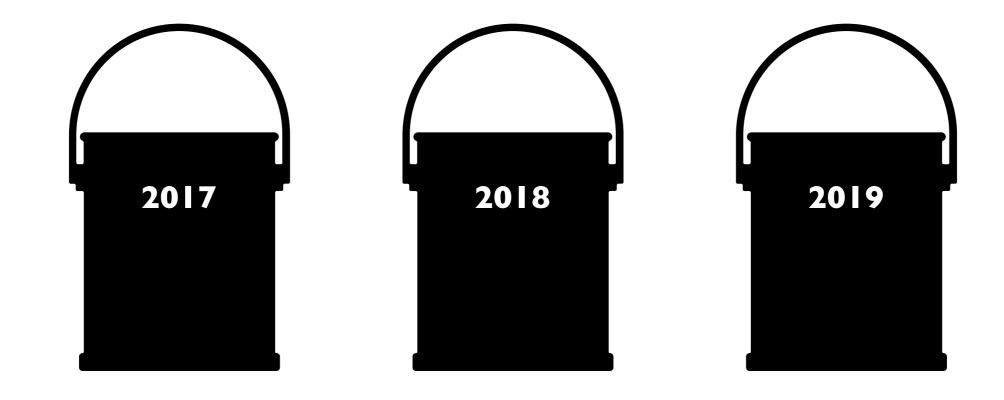
Probability Tables and Markov Chains (dict of dict)

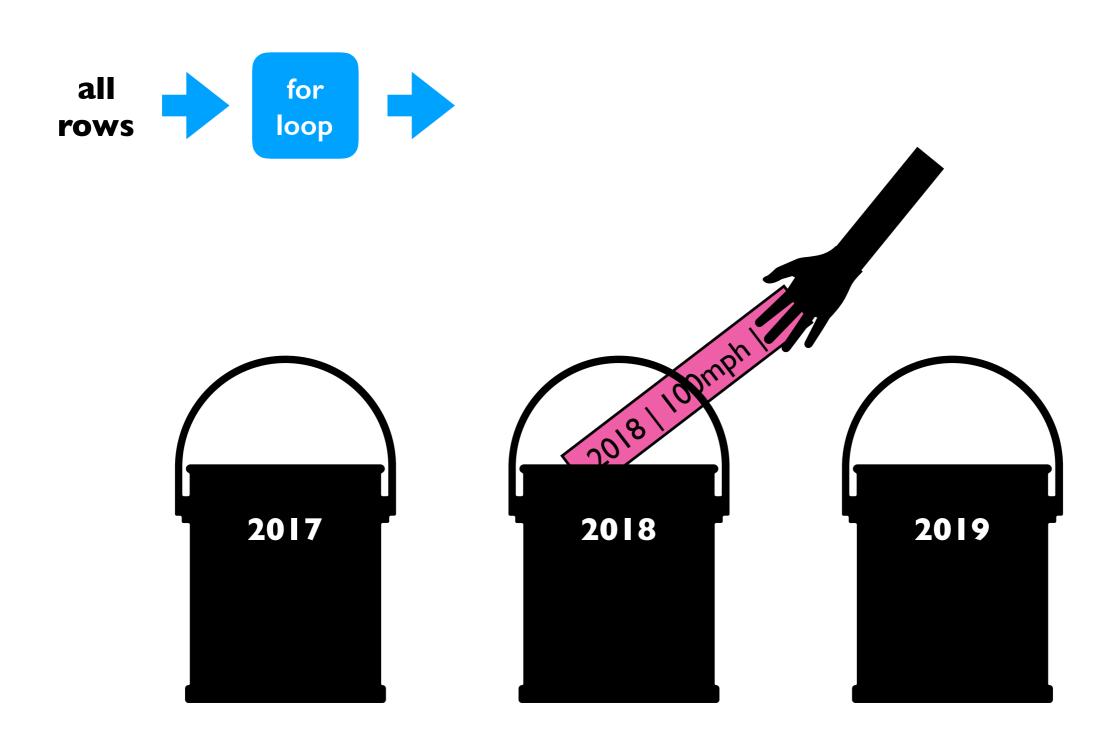




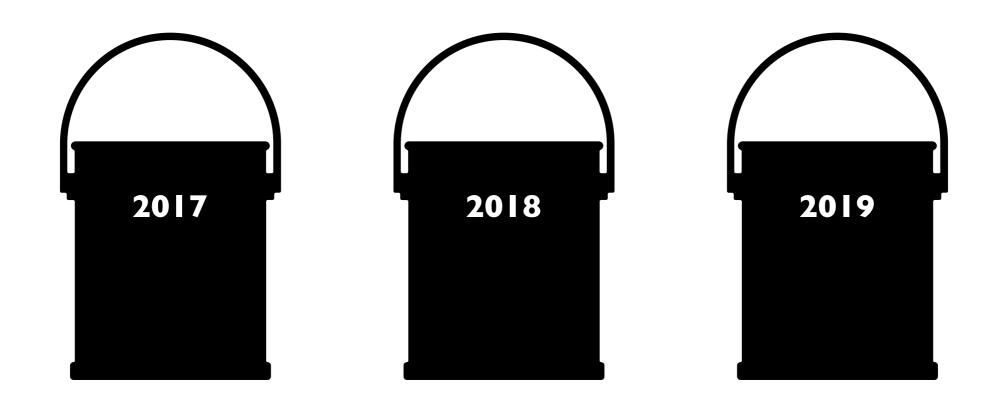




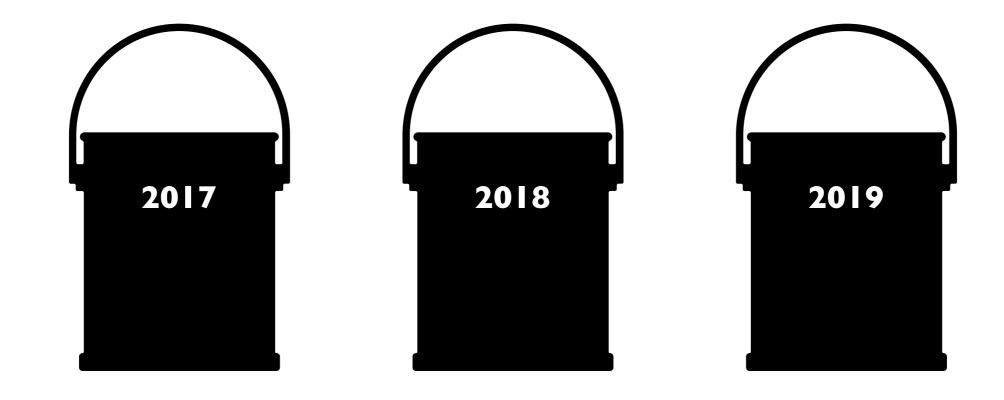


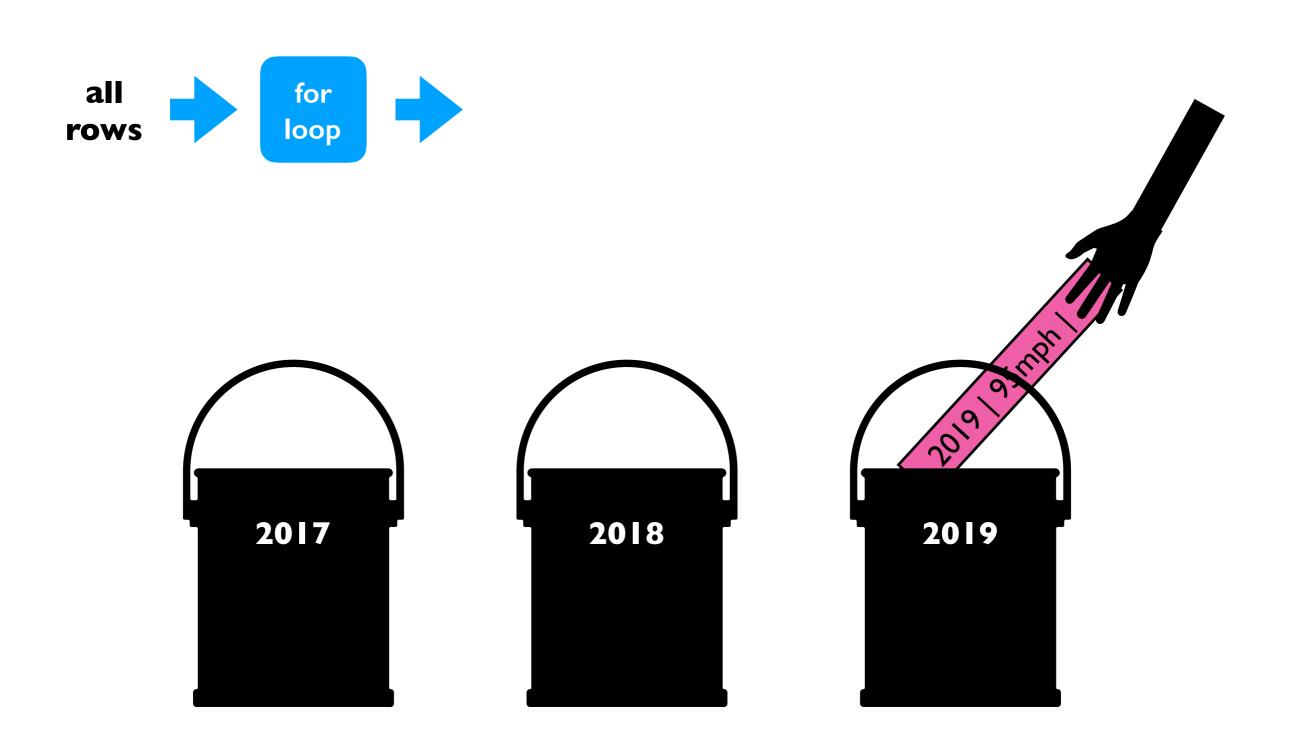


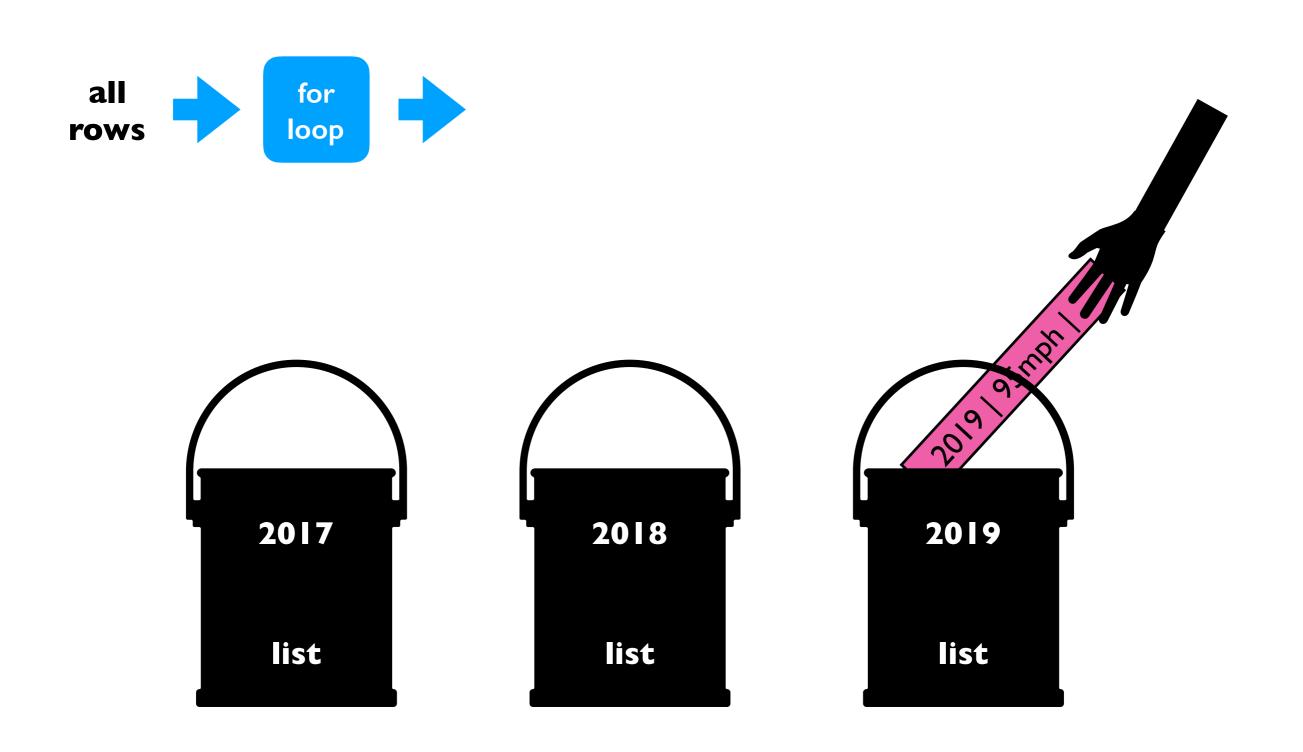


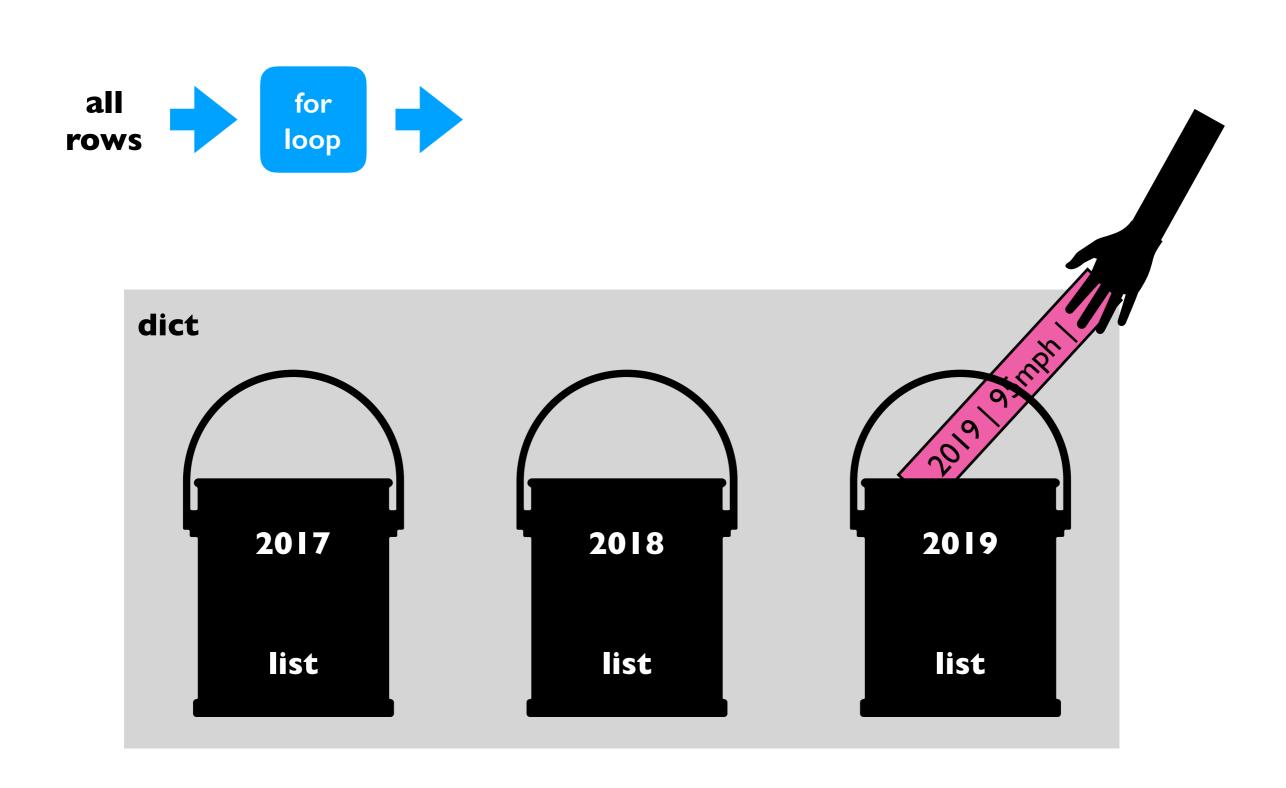












Bins with lists and dicts

all data

```
rows = [
    [2014, "A", 123],
    [2015, "B", 120],
    [2015, "C", 140],
    [2016, "D", 100],
    [2015, "E", 130],
    [2016, "F", 200],
]
```

Bins with lists and dicts

```
bins = \{
      all data
                                       2014: [
                                         [2014, "A", 123],
rows = [
                                       ],
                                       2015: [
  [2014, "A", 123],
  [2015, "B", 120],
                                         [2015, "B", 120],
  [2015, "C", 140],
                                         [2015, "C", 140],
  [2016, "D", 100],
                                         [2015, "E", 130],
  [2015, "E", 130],
                                       ],
  [2016, "F", 200],
                                       2016:
                                         [2016, "D", 100],
                                         [2016, "F", 200],
```

Bins with lists and dicts

```
bins = \{
      all data
                                       2014: [
                                                                      median 123
                                          [2014, "A", 123],
rows = [
                                       ],
  [2014, "A", 123],
                                       2015: [
  [2015, "B", 120],
                                          [2015, "B", 120],
                                                                      median 130
                                          [2015, "C", 140],
  [2015, "C", 140],
  [2016, "D", 100],
                                          [2015, "E", 130],
  [2015, "E", 130],
                                       ],
  [2016, "F", 200],
                                       2016: [
                                          [2016, "D", 100],
                                                                      median 150
                                          [2016, "F", 200],
```

Demo I: Median Tornado Speed per Year

Goal: print **median speed** of tornados for each year

Input:

Tornado CSV

Output:

Median within each year

Example:

prompt> python tornados.py

•••

2015: 130

2016: 123

2017:90

Today's Outline

Dictionary Ops

Binning (dict of list)

Table Representation (list of dict)

Probability Tables and Markov Chains (dict of dict)

Table Representation

name	X	у
Alice	30	20
Bob	5	11
Cindy	-2	50

list of list representation

```
header = ["name", "x", "y"]
rows = [
    ["Alice", 30, 20],
    ["Bob", 5, 11],
    ["Cindy", -2, 50],
]
```

Table Representation

name	X	y
Alice	30	20
Bob	5	11
Cindy	-2	50

list of list representation

list of dict representation

```
header = ["name", "x", "y"]

rows = [
    ["Alice", 30, 20],
    ["Bob", 5, 11],
    ["Cindy", -2, 50],
]
["cindy", -2, 50],
```

```
{"name":"Alice", "x":30, "y":20"},
{"name":"Bob", "x":5, "y":11"},
{"name":"Cindy", "x":-2, "y":50"},
]
```

Table Representation

name	x	у
Alice	30	20
Bob	5	11
Cindy	-2	50

list of list representation

list of dict representation

```
header = ["name", "x", "y"]
rows = [
     ["Alice", 30, 20],
     ["Bob", 5, 11],
2 → ["Cindy", -2, 50],
]
```

rows[2][header.index("y")]

rows[2]["y"]

Demo 2: Table Transform

Goal: create function that transforms list of lists table to a list of dicts table

Input:

• List of lists (from a CSV)

Output:

• List of dicts

Example:

```
>>> header = ["x";"y"]
>>> rows = [[1,2], [3,4]]
>>> transform(header, rows)
[{"x":1,"y":2}, {"x":3,"y":4}]
```

Today's Outline

Dictionary Ops

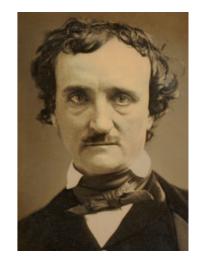
Binning (dict of list)

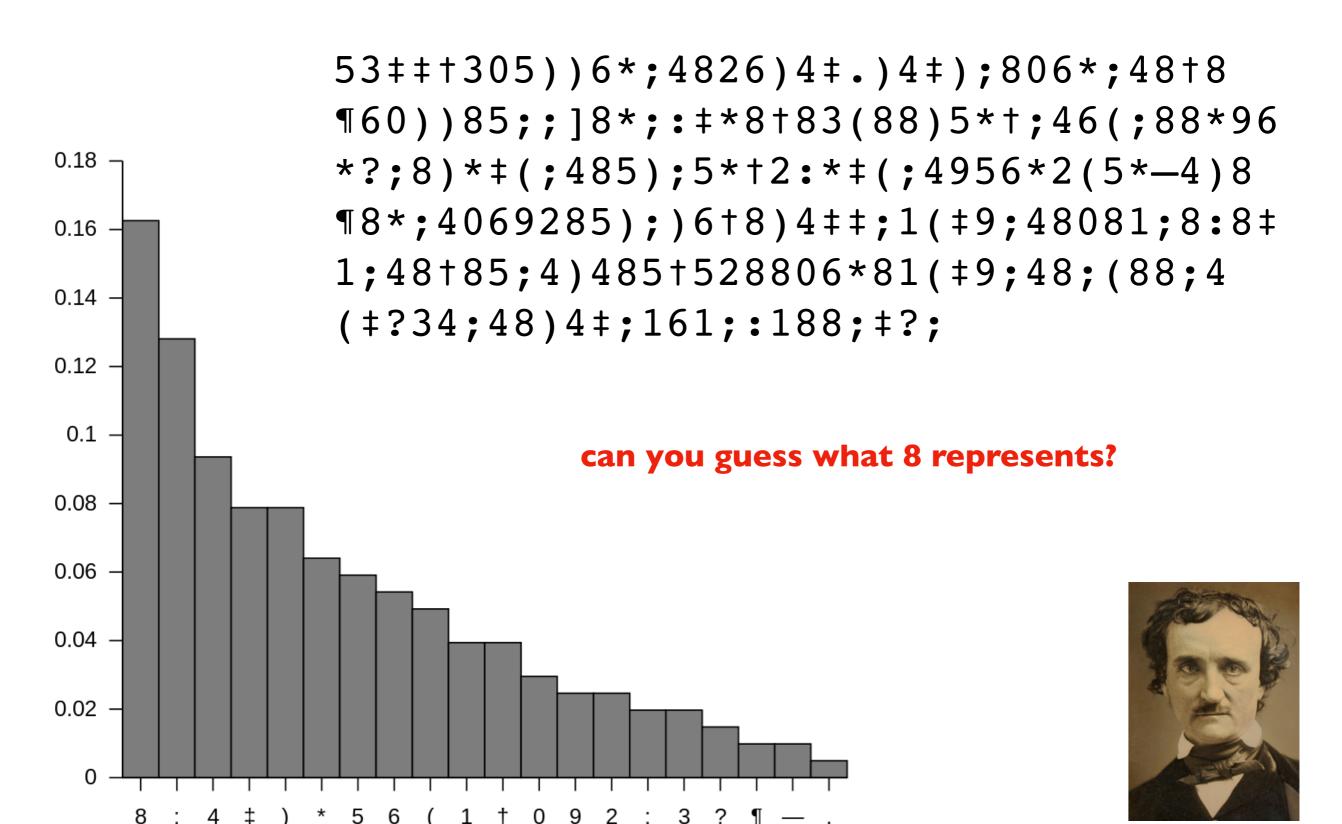
Table Representation (list of dict)

Probability Tables and Markov Chains (dict of dict)

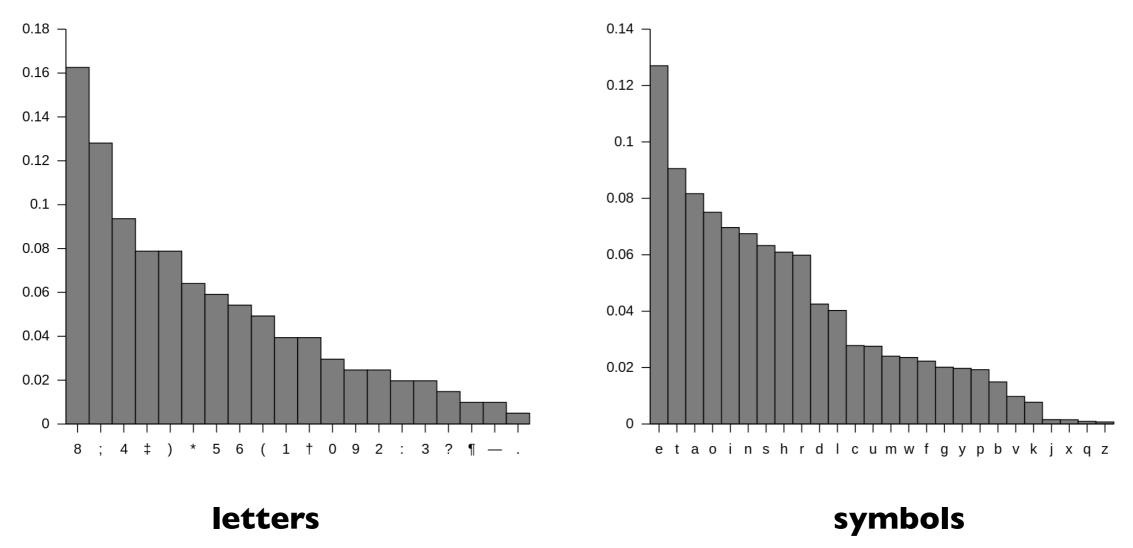
```
53‡‡†305))6*;4826)4‡.)4‡);806*;48†8
¶60))85;;]8*;:‡*8†83(88)5*†;46(;88*96
*?;8)*‡(;485);5*†2:*‡(;4956*2(5*—4)8
¶8*;4069285);)6†8)4‡‡;1(‡9;48081;8:8‡
1;48†85;4)485†528806*81(‡9;48;(88;4)(‡?34;48)4‡;161;:188;‡?;
```

can you guess what 8 represents?

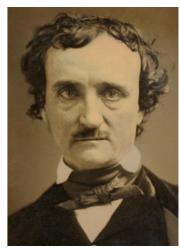




https://en.wikipedia.org/wiki/The_Gold-Bug



how to compute these?



https://en.wikipedia.org/wiki/The_Gold-Bug

Goal: if we randomly pick a word in a text, what is the probability that it will be a given letter?

Input:

Plaintext of book (from Project Gutenberg)

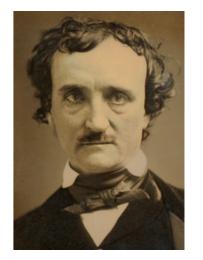
Output:

The portion of letters in the text that are that letter

Example:

text: AAAAABBCCC

A: 50% B: 20% C: 30%



https://en.wikipedia.org/wiki/The_Gold-Bug

Consider this sequence: "the quick tiger is quiet"

What letter likely comes after "t" in this text?

Consider this sequence: "the quick tiger is quiet"

What letter likely comes after "t" in this text?

N ext Letter	Probability	
h	50%	dict for "t":
i	50%	{"h": 0.5, "i": 0.5}
а	0%	
•••	0%	

Consider this sequence: "the quick tiger is quiet"

What letter likely comes after "t" in this text?

N ext Letter	Probability	
h	50%	dict for "t":
i	50%	{"h": 0.5, "i": 0.5}
а	0%	
•••	0%	

What letter likely comes after "q" in this text?

N ext Letter	Probability		
u	100%	dict for "q": {"u": 1.0}	
•••	0%	ι α • 1•0}	

Consider this sequence: "the quality limits and limits

What letter likely comes after "t" in this text?

	dict for "t":
	{"h": 0.5, "i": 0.5}

What letter likely comes after "q" in this text?

	dict for "q": {"u": 1.0}	
	(u • 1•0)	

Consider this sequence: "the quality limits and a next-letter probability dictionary for every letter

What letter likely comes after '

Next Letter Probability u 100% ... 0%

dict for "u":

{"i": 1.0}

dict for "t":

{"h": 0.5, "i": 0.5}

dict for "i":

{"c": 0.25, "g": 0.25,
"s": 0.25, "e": 0.25}

dict for "q":

{"u": 1.0}

Organize all the dicts with a dict:

Imagine a next-letter probability dictionary for every letter

```
dict for "u":
{"i": 1.0}

dict for "t":
{"h": 0.5, "i": 0.5}
```

dict for "i":

```
{"c": 0.25, "g": 0.25, "s": 0.25, "e": 0.25}
```

dict for "q": {"u": 1.0}



Organize all the dicts with a dict:

rgainze an the dicts with a dict

```
probs = {
  "u": {"i": 1.0},
                                       dict for "u":
                                       {"i": 1.0}
                                       dict for "t":
                                       {"h": 0.5, "i": 0.5}
}
                                       dict for "i":
                                       {"c": 0.25, "g": 0.25,
                                       "s": 0.25, "e": 0.25}
                                       dict for "q":
                                       {"u": 1.0}
```

Organize all the dicts with a dict:

```
dict for "u":
{"i": 1.0}
dict for "t":
{"h": 0.5, "i": 0.5}
dict for "i":
{"c": 0.25, "g": 0.25,
"s": 0.25, "e": 0.25}
dict for "q":
{"u": 1.0}
```

Organize all the dicts with a dict:

```
probs["i"]
```

```
dict for "u":
{"i": 1.0}
dict for "t":
{"h": 0.5, "i": 0.5}
dict for "i":
{"c": 0.25, "g": 0.25,
"s": 0.25, "e": 0.25}
dict for "q":
{"u": 1.0}
```

Organize all the dicts with a dict:

```
probs["i"]["e"] = 0.25
```

There is a 25% probability that the letter following an "i" is an "e"

```
dict for "u":
{"i": 1.0}
dict for "t":
{"h": 0.5, "i": 0.5}
dict for "i":
{"c": 0.25, "g": 0.25,
"s": 0.25, "e": 0.25}
dict for "q":
{"u": 1.0}
```

Vocabulary

The collection of transition probabilities like this is sometimes called a

"stochastic matrix"

Processes that make probabilistic transitions like this (e.g., from one letter to the next) are called "Markov chains"

Random Text Generation

ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.

which looks closest to English?

OCRO HLI RGWR NMIELWIS EU LL

NBNESEBYA TH EEI ALHENHTTPA

OOBTTVA NAH BRL.

INCTORE ST BE S DEAMY ACHIN D
ILONASIVE TUCOOWE AT
TEASONARE FUSO TIZIN ANDY
TOBE SEACE CTISBE.

ON IE ANTSOUTINYS ARE T

Examples from A Mind at Play, by Soni and Goodman

Random Text Generation

all letters equally likely

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.

weighted random, based on frequency in a text (implement with dict)

OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.

probability of each letter based on previous letter (implement with dict of dicts)

ON IE ANTSOUTINYS ARE T
INCTORE ST BE S DEAMY ACHIN D
ILONASIVE TUCOOWE AT
TEASONARE FUSO TIZIN ANDY
TOBE SEACE CTISBE.

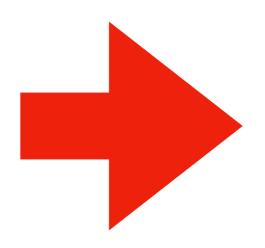
Hypothetical Use Case

DNA sequences

GATACAGATACA

GCTATAGCTATAGCGCGC

AAAATTTTAAAATTTTTAAAA



stochastic model



BIOINFORMATICS APPLICATIONS NOTE Vol. 22 no. 12 2006, pages 1534–1535 doi:10.1093/bioinformatics/btl113

Sequence analysis

GenRGenS: software for generating random genomic sequences and structures

Yann Ponty¹, Michel Termier² and Alain Denise^{1,*}

¹LRI, UMR CNRS 8623, Université Paris-Sud 11, F91405 Orsay cedex, France and ²IGM, UMR CNRS 8621, Université Paris-Sud 11, F91405 Orsay cedex, France

Received on February 21, 2006; revised on March 13, 2006; accepted on March 21, 2006

Advance Access publication March 30, 2006

Associate Editor: Martin Bishop

CATCATC?TC?TCATC?TCAT CATCATCATCATCATCAT

synthetic sequences, filling in gaps

Demo 4: Conditional Letter Frequency

Goal: if we look at given letter, what is the next letter likely to be?

Input:

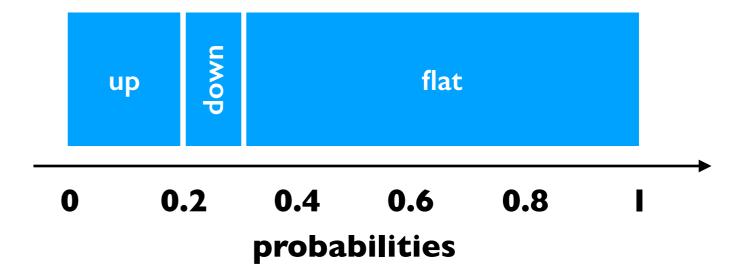
Plaintext of book (from Project Gutenberg)

Output:

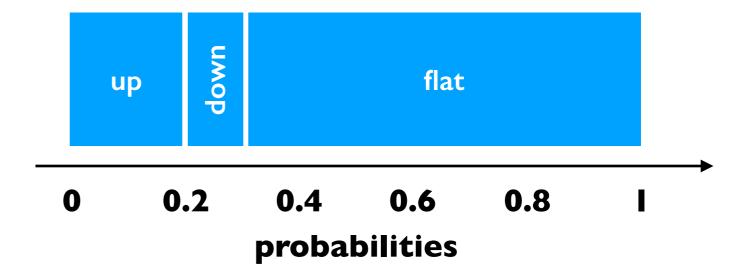
- Transition probabilities
- Randomly generated text, based on probabilities

```
transitions = {
    "up": 0.2,
    "down": 0.1,
    "flat": 0.7
}
```

```
transitions = {
    "up": 0.2,
    "down": 0.1,
    "flat": 0.7
}
```

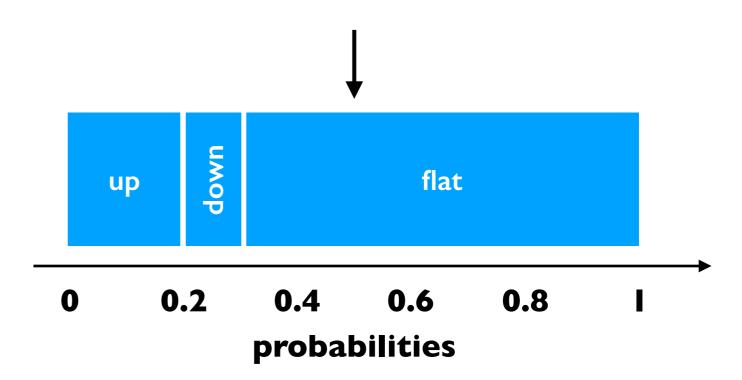


```
transitions = {
    "up": 0.2,
    "down": 0.1,
    "flat": 0.7
}
```



```
transitions = {
    "up": 0.2,
    "down": 0.1,
    "flat": 0.7
}

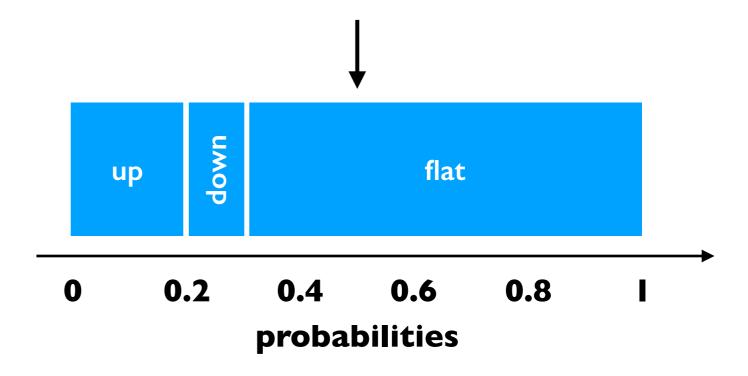
x = random.random()
# assume 0.5
```



```
transitions = {
    "up": 0.2,
    "down": 0.1,
    "flat": 0.7
}

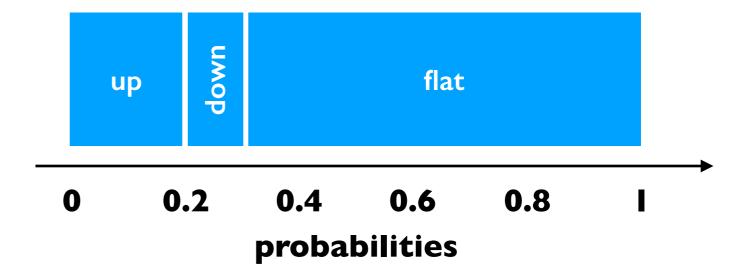
x = random.random()
# assume 0.5
```

flat "wins"



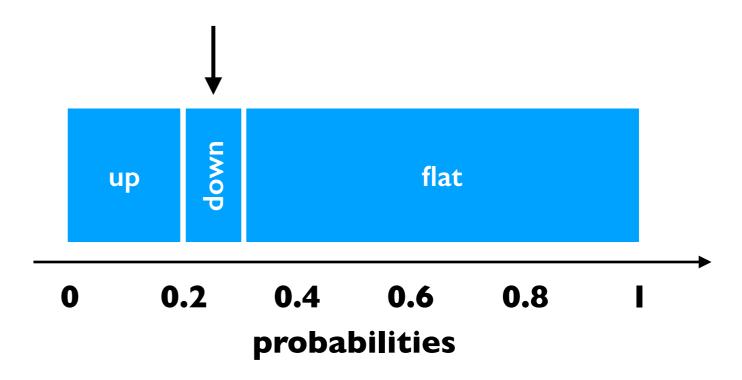
```
transitions = {
    "up": 0.2,
    "down": 0.1,
    "flat": 0.7
}

x = random.random()
# assume 0.25
```



```
transitions = {
    "up": 0.2,
    "down": 0.1,
    "flat": 0.7
}

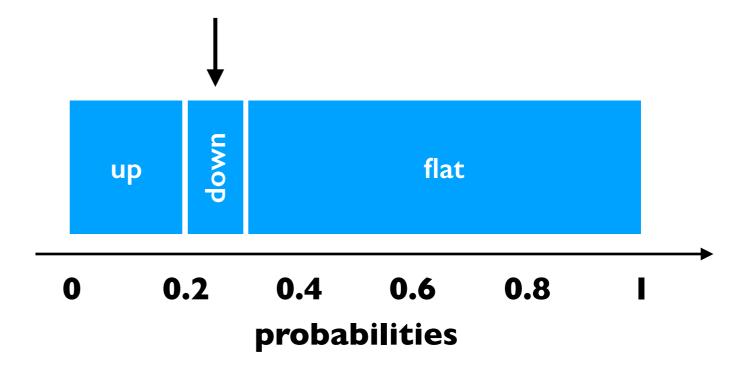
x = random.random()
# assume 0.25
```



```
transitions = {
    "up": 0.2,
    "down": 0.1,
    "flat": 0.7
}

x = random.random()
# assume 0.25
```

down "wins"



```
transitions = {
    "up": 0.2,
    "down": 0.1,
    "flat": 0.7
x = random.random()
\# assume 0.25
end = 0
keys = ["up", "down", "flat"]
winner = None
for key in keys:
    end += transitions[key]
    if end >= x:
        winner = key
        break
```

```
end
transitions = {
    "up": 0.2,
    "down": 0.1,
                                                   flat
    "flat": 0.7
                                  up
                                           0.4 0.6
                                     0.2
                                                       8.0
                                0
x = random.random()
                                          probabilities
\# assume 0.25
end = 0
keys = ["up", "down", "flat"]
winner = None
for key in keys:
                                         key up
   end += transitions[key]
    if end >= x:
                                         end 0
        winner = key
        break
```

```
end
transitions = {
    "up": 0.2,
    "down": 0.1,
                                                   flat
    "flat": 0.7
                                  up
                                           0.4 0.6
                                                       8.0
                                     0.2
                                0
x = random.random()
                                          probabilities
# assume 0.25
end = 0
keys = ["up", "down", "flat"]
winner = None
for key in keys:
                                         key up
    end += transitions[key]
    if end >= x:
                                         end 0.2
        winner = key
        break
```

```
end
transitions = {
    "up": 0.2,
    "down": 0.1,
                                                   flat
    "flat": 0.7
                                  up
                                           0.4 0.6
                                     0.2
                                                       8.0
                                0
x = random.random()
                                          probabilities
# assume 0.25
end = 0
keys = ["up", "down", "flat"]
winner = None
for key in keys:
                                         key up
    end += transitions[key]
    if end >= x:
                                         end 0.2
        winner = key
        break
```

```
end
transitions = {
    "up": 0.2,
    "down": 0.1,
                                                   flat
    "flat": 0.7
                                  up
                                           0.4 0.6
                                     0.2
                                                       8.0
                                0
x = random.random()
                                          probabilities
\# assume 0.25
end = 0
keys = ["up", "down", "flat"]
winner = None
for key in keys:
                                         key down
   end += transitions[key]
    if end >= x:
                                         end 0.2
        winner = key
        break
```

```
end
transitions = {
    "up": 0.2,
    "down": 0.1,
                                                   flat
    "flat": 0.7
                                  up
                                           0.4 0.6
                                     0.2
                                                       8.0
                                0
x = random.random()
                                          probabilities
# assume 0.25
end = 0
keys = ["up", "down", "flat"]
winner = None
for key in keys:
                                         key down
    end += transitions[key]
    if end >= x:
                                         end 0.3
        winner = key
        break
```

```
end
transitions = {
    "up": 0.2,
    "down": 0.1,
                                        down
    "flat": 0.7
                                                   flat
                                  up
                                     0.2
                                           0.4 0.6
                                                        8.0
                                0
x = random.random()
                                          probabilities
\# assume 0.25
end = 0
keys = ["up", "down", "flat"]
winner = None
for key in keys:
                                          key down
    end += transitions[key]
    if end >= x:
                                          end 0.3
       winner = key
        break
```

we randomly chose "down"