

File Reading

Chris Gregg

Based on Slides by Chris Piech and Mehran Sahami
CS106A, Stanford University

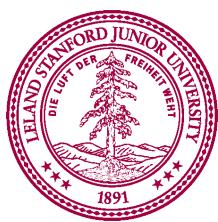
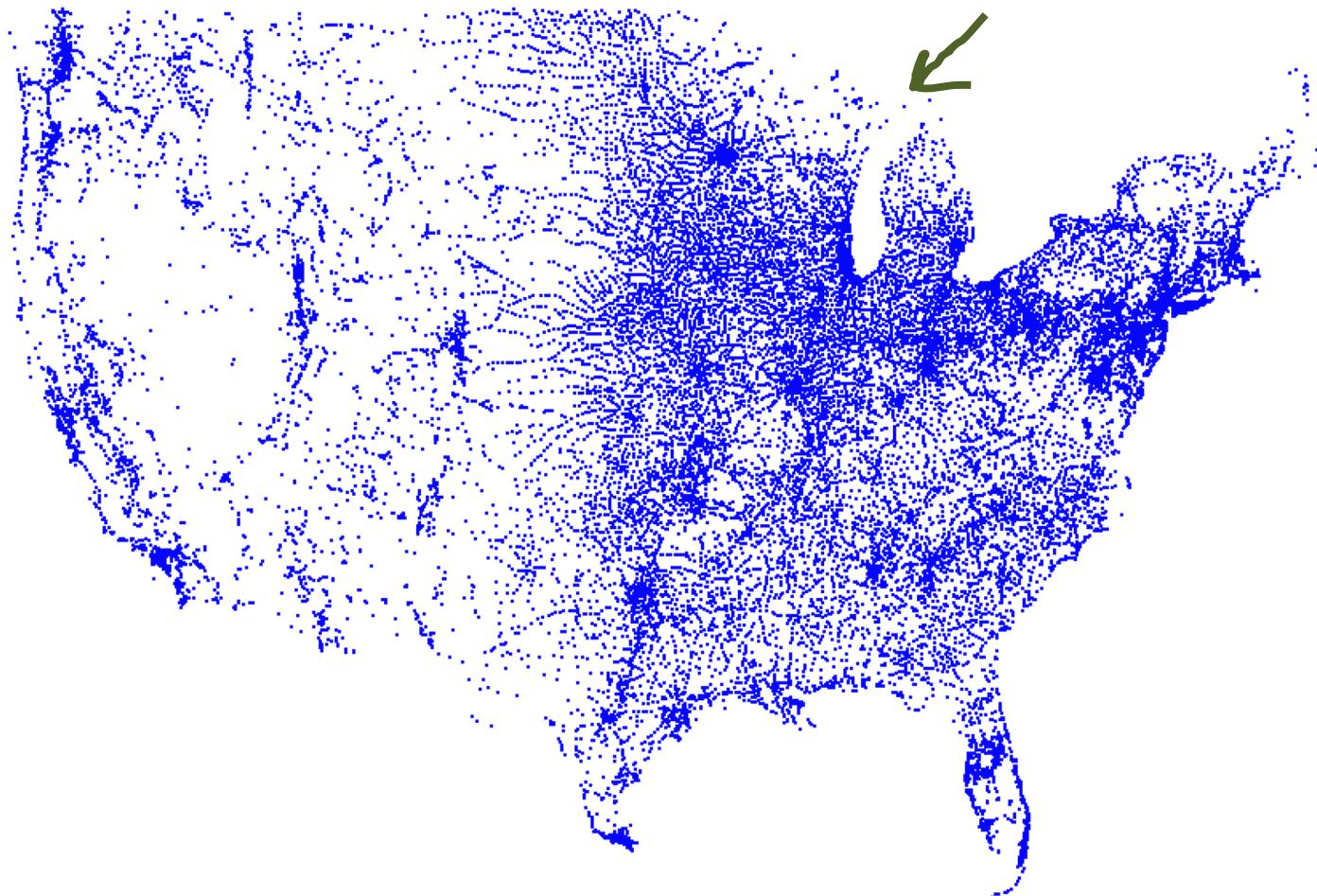
Learning Goals

1. Know how to read a file line by line.





Each blue dot is a tiny rectangle



Review

Strings are Immutable

```
x = 'hello there'
```



```
x[1] = 'z'
```

```
x = x[:1] + 'z' + x[2:]
```



End Review



ATGCTTAAACC..

Human Genome Project

"ATGCCAGGAC"

"GGACTTACATTTTT"

"ATTTTTGGCCGGCC"

The human genome has 3 billion base pairs



Compose Problem

strand1 "GCATCAT"

strand2 "CATTAC"

result "GCATCATTAC"

The human genome has 3 billion base pairs



Ha. Gene was working on
The Genome project ☺



Did Gene Myers define all those
little pieces as constants?

File Processing!

Getting Data into Programs

- Put it directly in the program:
 - Define constants holding your values.
- Get it from the user:
 - Mouse location, input, etc.
- Generate it randomly:
 - Use random.
- Get it from an external source.
 - Store it in a file and read it later.



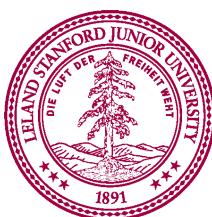
Reading Files

- Virtually all programs that you've used at some point read files from disk:
 - Word processing (documents)
 - Web browser (cookies)
 - Games (saved progress)
 - PyCharm (Python files)
 - Music player (songs)



The structure of files

- A file is just a series of **bits** (ones and zeros).
- Those bits can have structure:
 - Plain-text: Bits represent characters.
 - JPEG: Bits encode information about the structure of an image.
 - MP3: Bits encode frequency information about music.
 - etc.



The structure of files

A file is just a series of *bits* (ones and zeros).

Those bits can have structure:

- Plain-text: Bits represent characters.

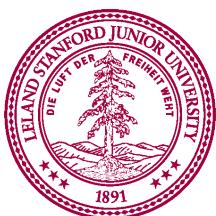
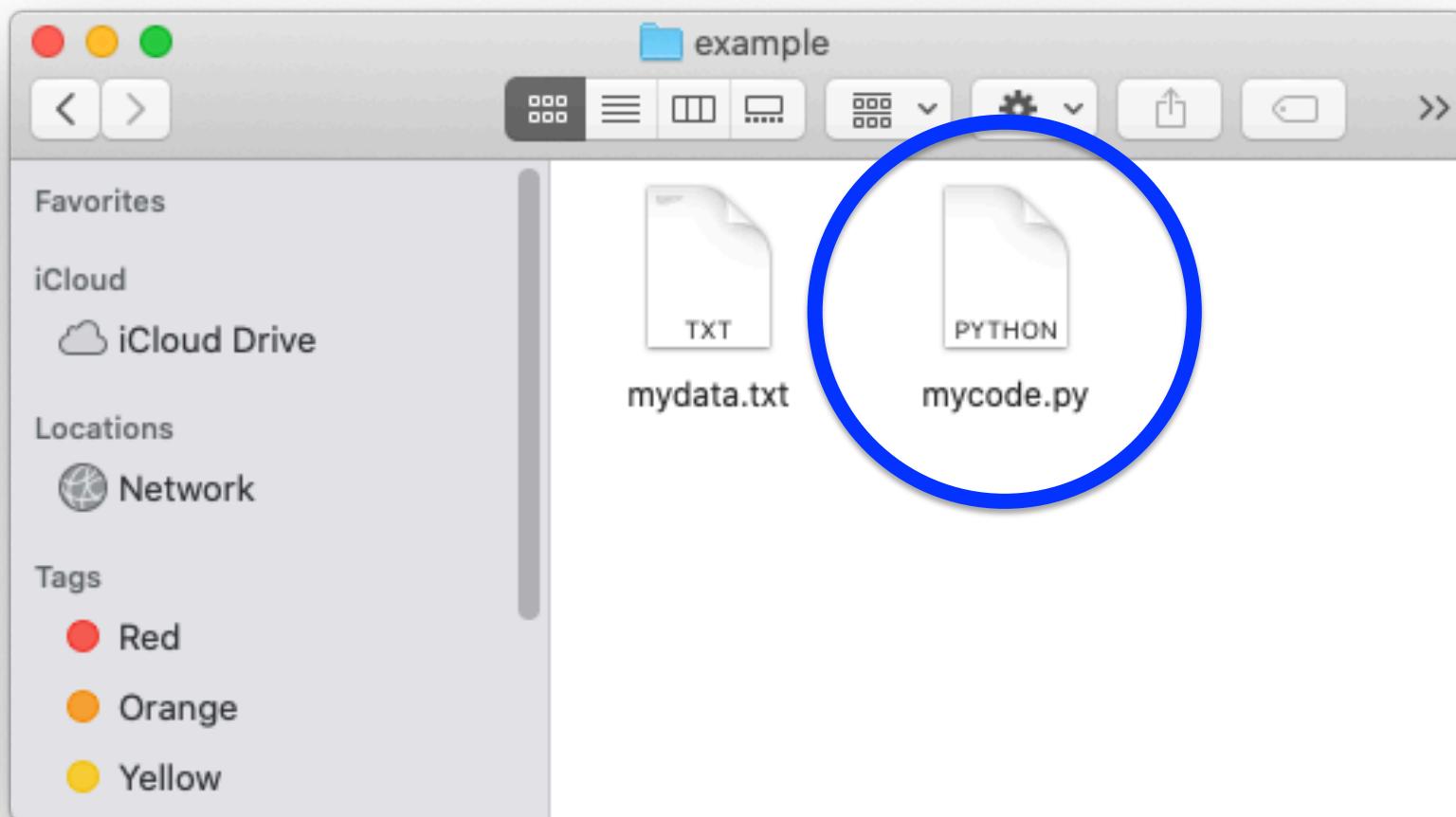
JPEG: Bits encode information about the structure of an image.

MP3: Bits encode frequency information about music.

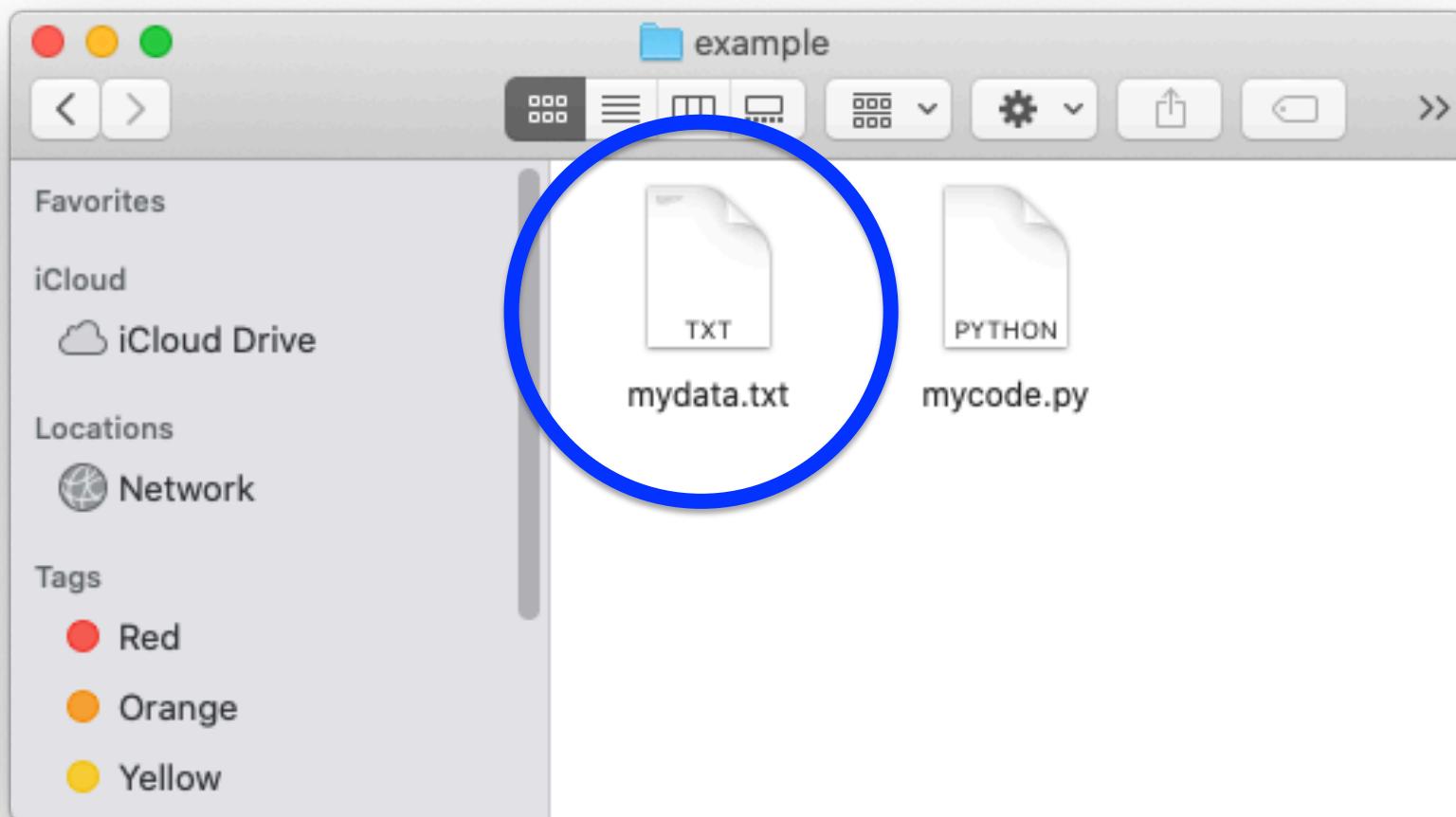
etc.



Whats a file?

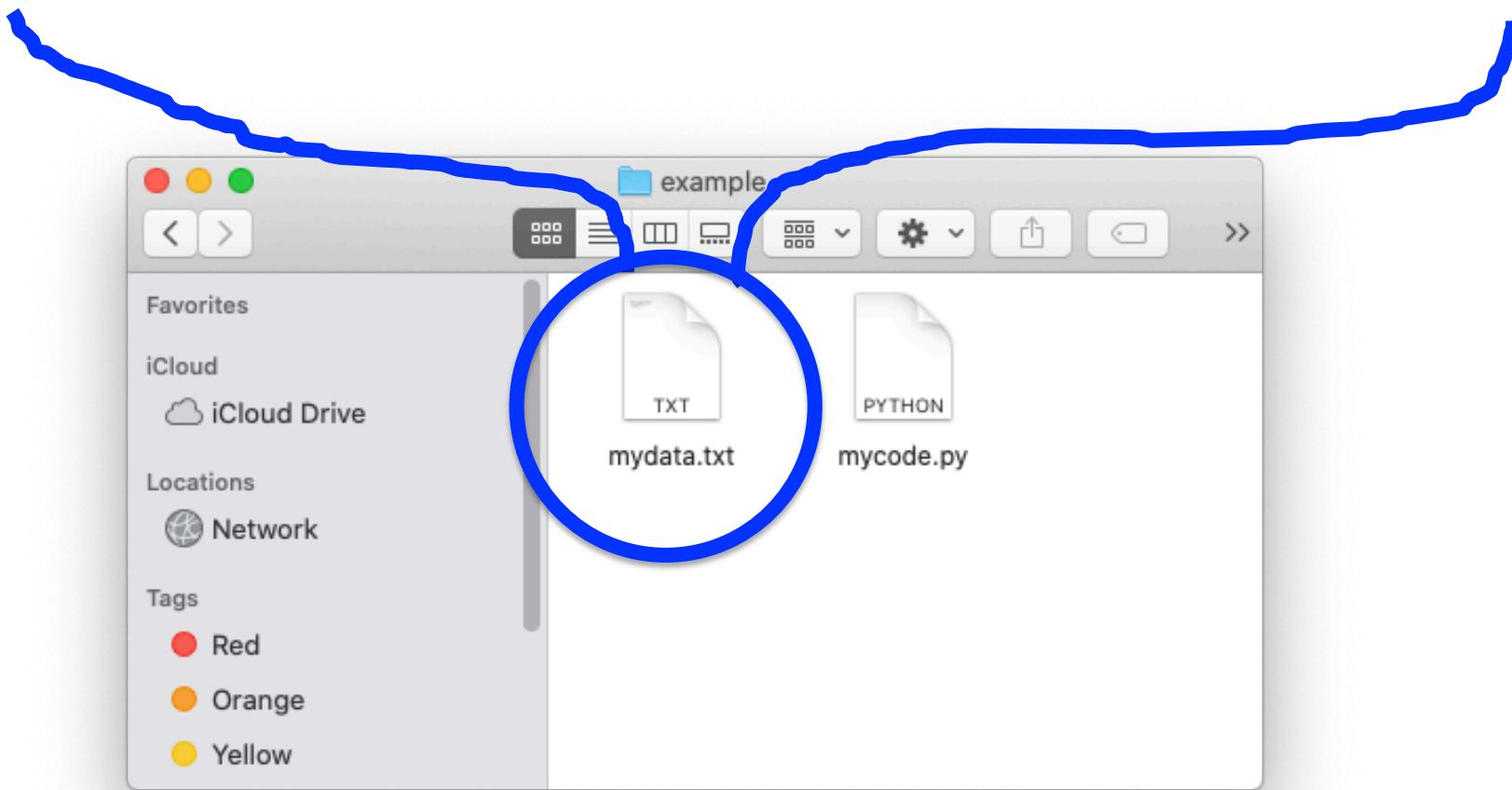


Whats a file?



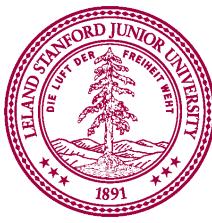
Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

– Max Ehrmann "Desiderata"



**Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.**

- Max Ehrmann "Desiderata"

```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

**Step One:
Open the file**



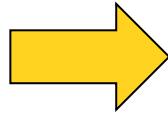
**Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.**

- Max Ehrmann "Desiderata"

```
file = open('mydata.txt')
for line in file:
    print(line)
file.close()
```

**Step One:
Open the file**





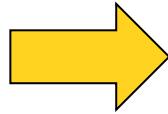
Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"

```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

Step Two:
Read the file one line at a time



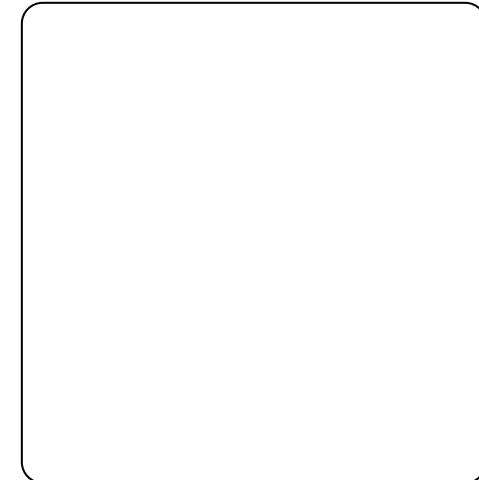


Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"

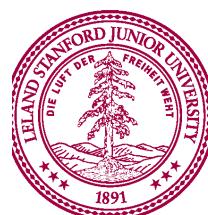
console

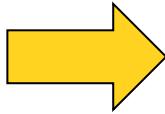
```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```



Step Two:

Read the file one line at a time





Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"

console

```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

Neither be cynical about love;

Step Two:
Read the file one line at a time



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

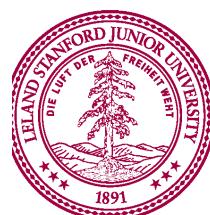
- Max Ehrmann "Desiderata"

```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

console

```
Neither be cynical about love;
```

Step Two:
Read the file one line at a time



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

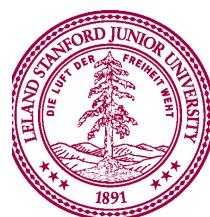
- Max Ehrmann "Desiderata"

console

```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

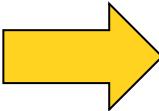
```
Neither be cynical about love;  
  
for in the face of all aridity
```

Step Two:
Read the file one line at a time



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"



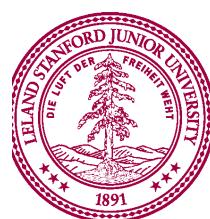
```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

console

```
Neither be cynical about love;  
  
for in the face of all aridity
```

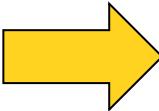
Step Two:

Read the file one line at a time



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"

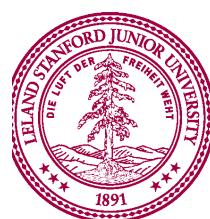


```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

console

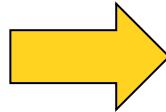
```
Neither be cynical about love;  
  
for in the face of all aridity  
  
and disenchantment it is as
```

Step Two:
Read the file one line at a time



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"



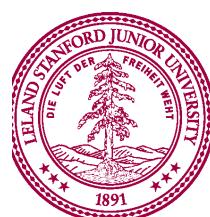
console

```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

```
Neither be cynical about love;  
  
for in the face of all aridity  
  
and disenchantment it is as
```

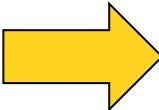
Step Two:

Read the file one line at a time



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"



```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

console

```
Neither be cynical about love;  
  
for in the face of all aridity  
  
and disenchantment it is as  
  
perennial as the grass.
```

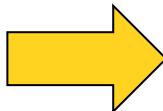
Step Two:

Read the file one line at a time



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

– Max Ehrmann "Desiderata"

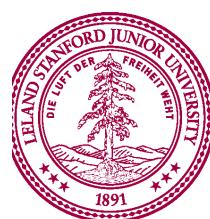


```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

console

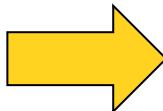
```
Neither be cynical about love;  
  
for in the face of all aridity  
  
and disenchantment it is as  
  
perennial as the grass.
```

Step Two:
Read the file one line at a time



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

– Max Ehrmann "Desiderata"

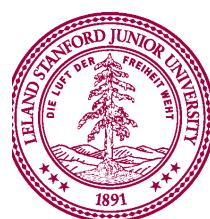


```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

console

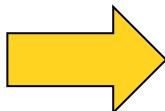
```
Neither be cynical about love;  
  
for in the face of all aridity  
  
and disenchantment it is as  
  
perennial as the grass.  
  
– Max Ehrmann "Desiderata"
```

Step Two:
Read the file one line at a time



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"



```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

console

```
Neither be cynical about love;  
  
for in the face of all aridity  
  
and disenchantment it is as  
  
perennial as the grass.  
  
- Max Ehrmann "Desiderata"
```

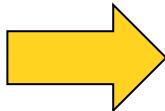
Step Two:

Read the file one line at a time



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"

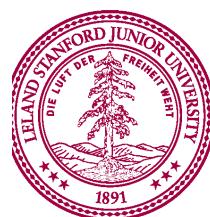


```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

console

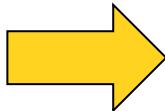
```
Neither be cynical about love;  
  
for in the face of all aridity  
  
and disenchantment it is as  
  
perennial as the grass.  
  
- Max Ehrmann "Desiderata"
```

Step Three: Close the file

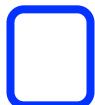


Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"



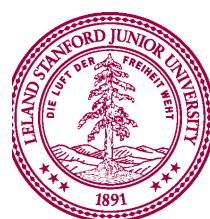
```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```



Step Three:
Close the file

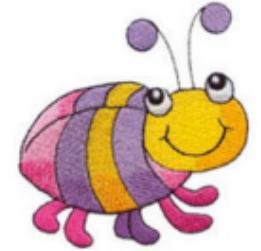
console

```
Neither be cynical about love;  
  
for in the face of all aridity  
  
and disenchantment it is as  
  
perennial as the grass.  
  
- Max Ehrmann "Desiderata"
```



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"



```
file = open('mydata.txt')  
for line in file:  
    print(line)  
file.close()
```

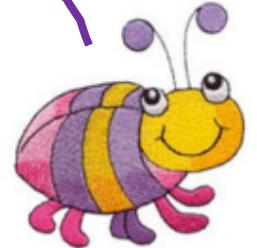
console

```
Neither be cynical about love;  
  
for in the face of all aridity  
  
and disenchantment it is as  
  
perennial as the grass.  
  
- Max Ehrmann "Desiderata"
```



line

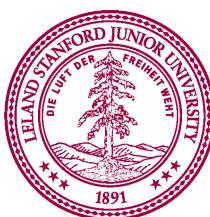
'perennial as the grass.\n'



```
file = open('mydata.txt')
for line in file:
    print(line)
file.close()
```

console

Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.



Take 2

line

'perennial as the grass.'

```
file = open('mydata.txt')
for line in file:
    line = line[:-1]
    print(line)
file.close()
```



Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.

- Max Ehrmann "Desiderata"

console

```
file = open('mydata.txt')  
for line in file:  
    line = line[:-1]  
    print(line)  
file.close()
```

Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass.
- Max Ehrmann "Desiderata"





What does this do?

```
f = open('mydata.txt')  
for line in f:  
    print(line[:-1])  
print('-----')  
for line in f:  
    print(line[:-1])
```

Option A

Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass
- Max Ehrmann "Desiderata"

Option B

Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass
- Max Ehrmann "Desiderata"

Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass
- Max Ehrmann "Desiderata"

What does this do?

```
f = open('mydata.txt')  
for line in f:  
    print(line[:-1])  
print('-----')  
for line in f:  
    print(line[:-1])
```

The file object
doesn't reset after
the for loop

Don't read from a
file twice...

Option A

Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass
- Max Ehrmann "Desiderata"

Option B

Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass
- Max Ehrmann "Desiderata"

Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass
- Max Ehrmann "Desiderata"

Round 2

What does this do?

```
f = open('mydata.txt')  
next(f)  
for line in f:  
    print(line[:-1])
```

Option A

Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass
- Max Ehrmann "Desiderata"

Option B

for in the face of all aridity
and disenchantment it is as
perennial as the grass
- Max Ehrmann "Desiderata"



What does this do?

next skips
a line (and
returns it)

```
f = open('mydata.txt')  
next(f)  
for line in f:  
    print(line[:-1])
```



Option A

Neither be cynical about love;
for in the face of all aridity
and disenchantment it is as
perennial as the grass
- Max Ehrmann "Desiderata"

Option B

for in the face of all aridity
and disenchantment it is as
perennial as the grass
- Max Ehrmann "Desiderata"







with or without you?

How edgy are you feeling?

This is the proper way to do it...

```
with open('mydata.txt') as f:  
    for line in file:  
        line = line[:-1]  
    print(line)
```



with recruits the "file manager" to help python know when it can "close" the file



How edgy are you feeling?

```
for line in open('mydata.txt'):  
    line = line[:-1]  
    print(line)
```

You actually don't need
with. The garbage
collector will close your
file for you...



How edgy are you feeling?

```
with open('mydata.txt') as f:  
    for line in f:  
        line = line[:-1]  
    print(line)
```



But some old versions of Python don't garbage collect and then the file is open till the program ends wasting resources

Semisonic wrote a song called "Closing Time"



I'm a little edgy when it comes to files

```
for line in open('mydata.txt'):  
    line = line[:-1]  
    print(line)
```

The 90s called.
They want python2 back...

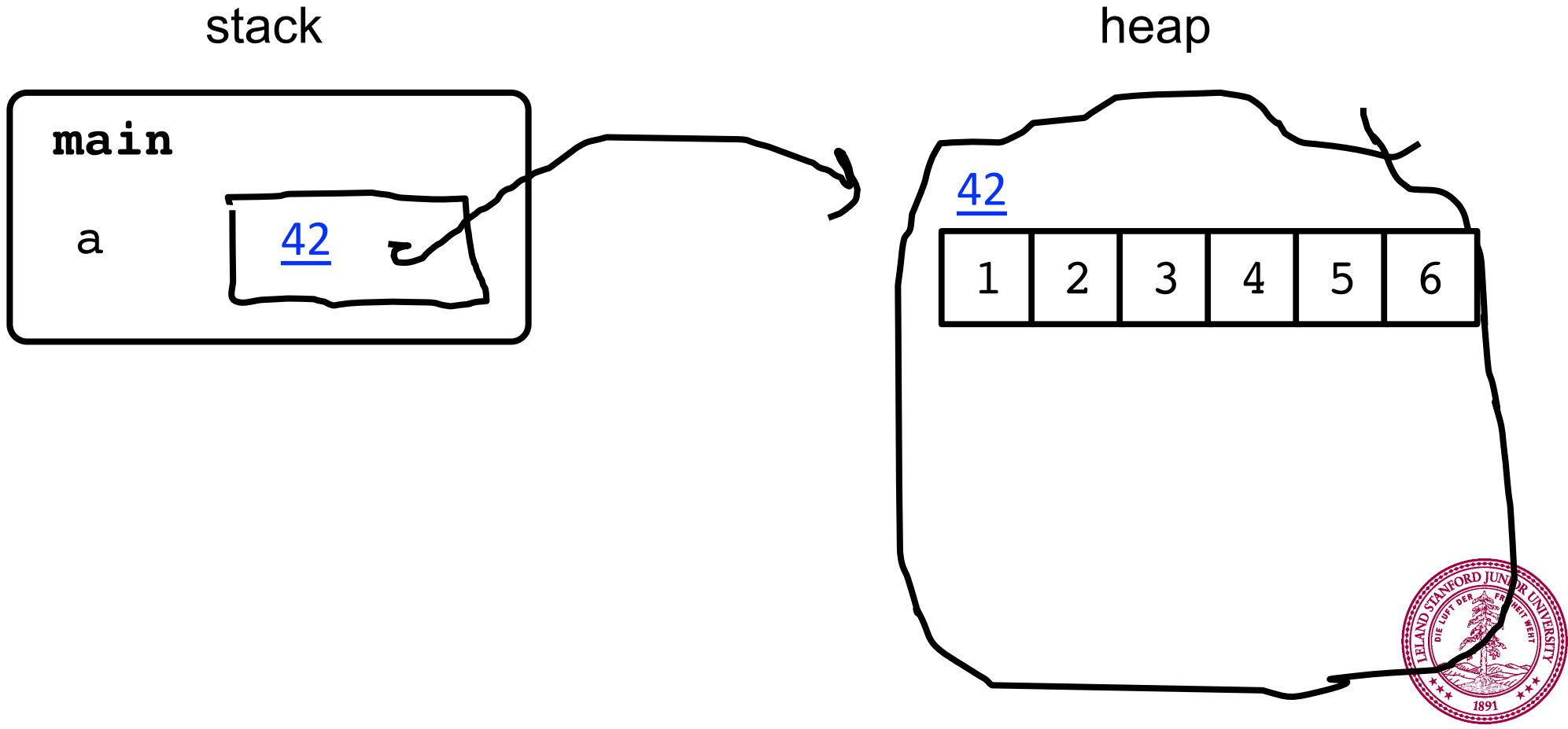


Aside: What is the Garbage Collector?



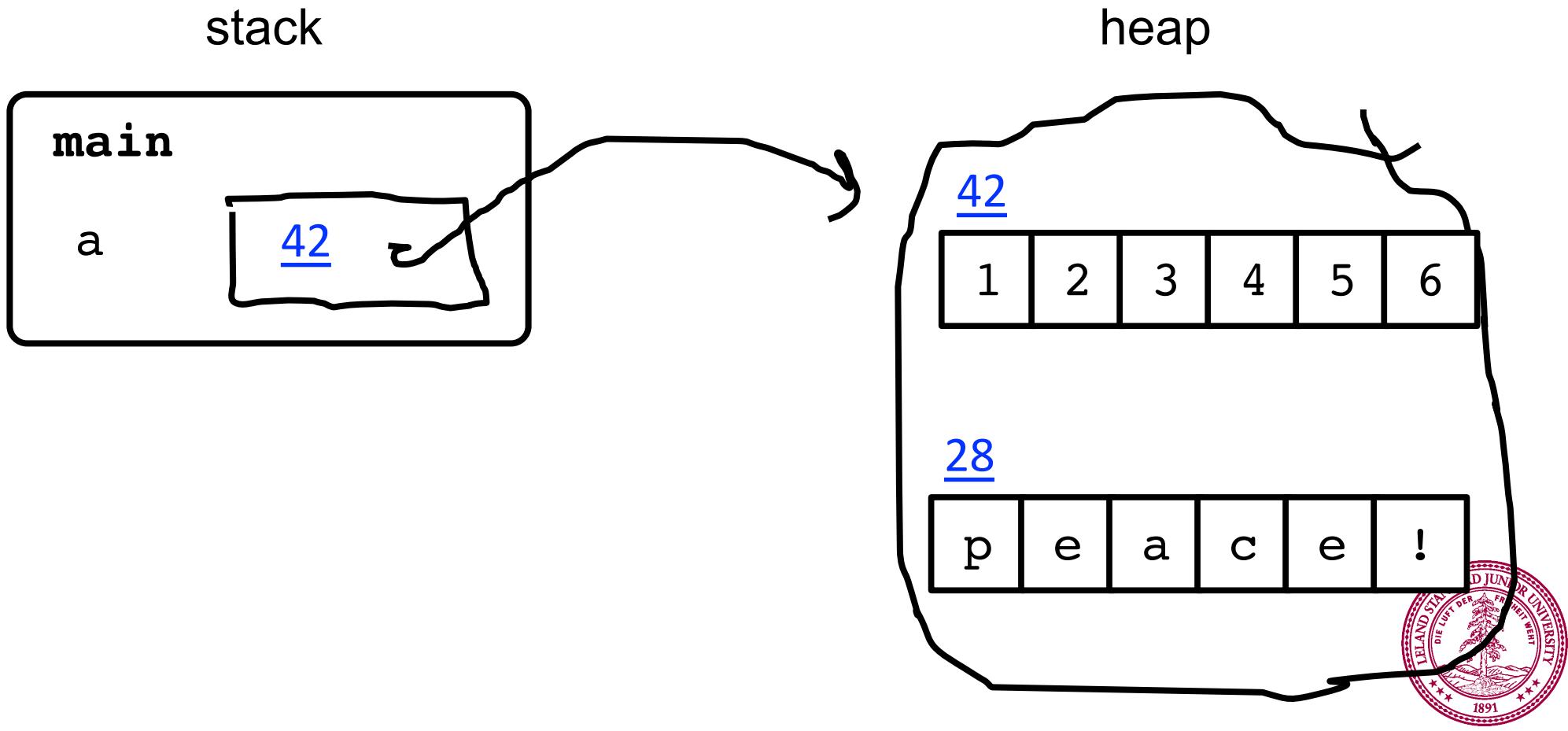
Aside: What is the Garbage Collector?

```
a = [1, 2, 3, 4, 5, 6]  
a = peace!
```



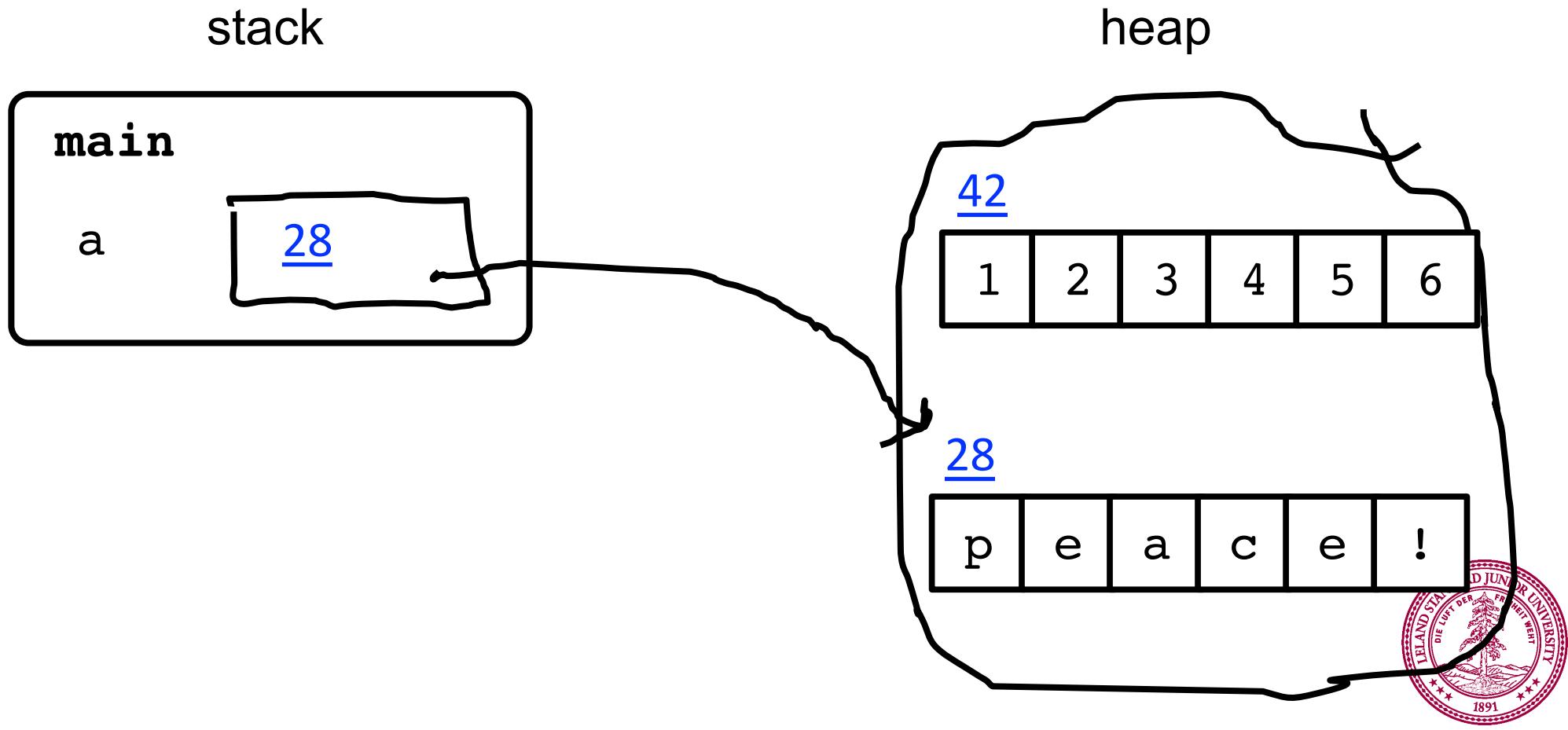
Aside: What is the Garbage Collector?

```
a = [1, 2, 3, 4, 5, 6]  
a = 'peace! '
```



Aside: What is the Garbage Collector?

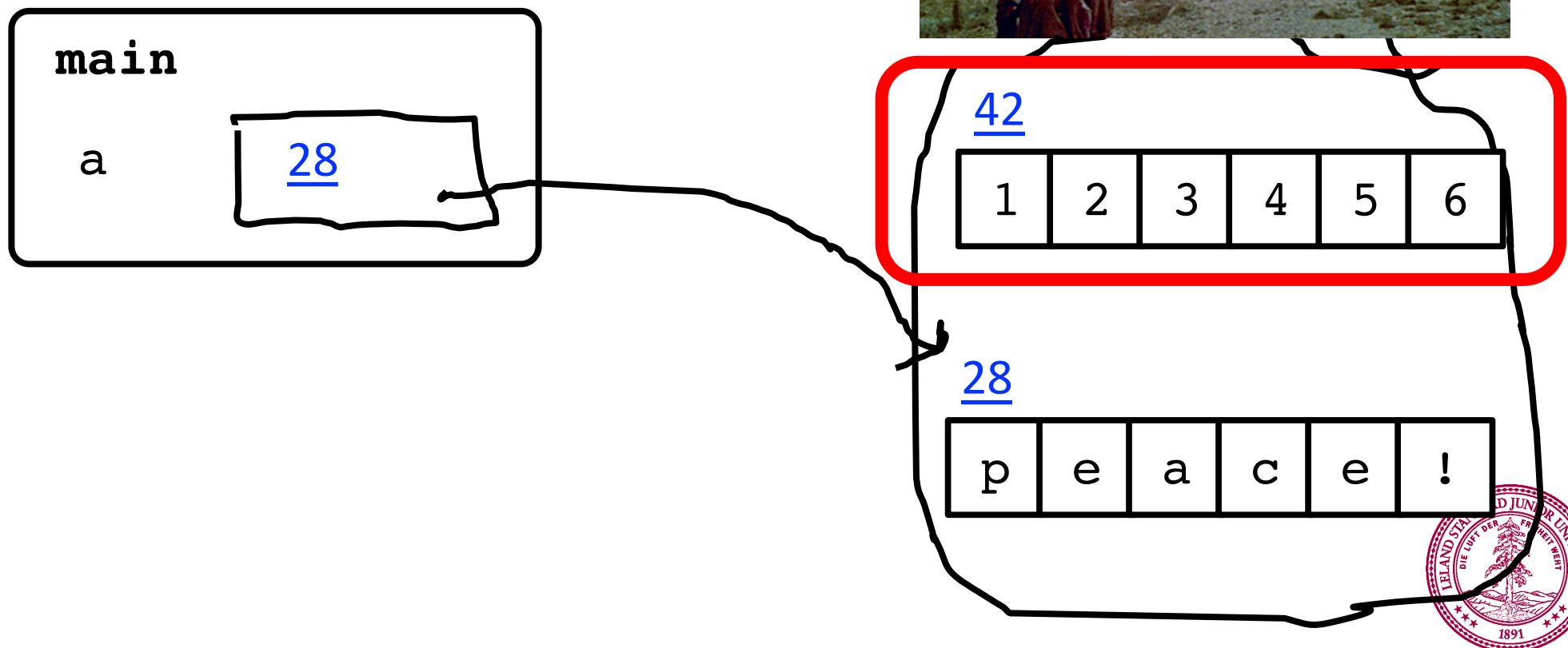
```
a = [1, 2, 3, 4, 5, 6]  
a = 'peace!'
```



Aside: What is the Garbage Collector?

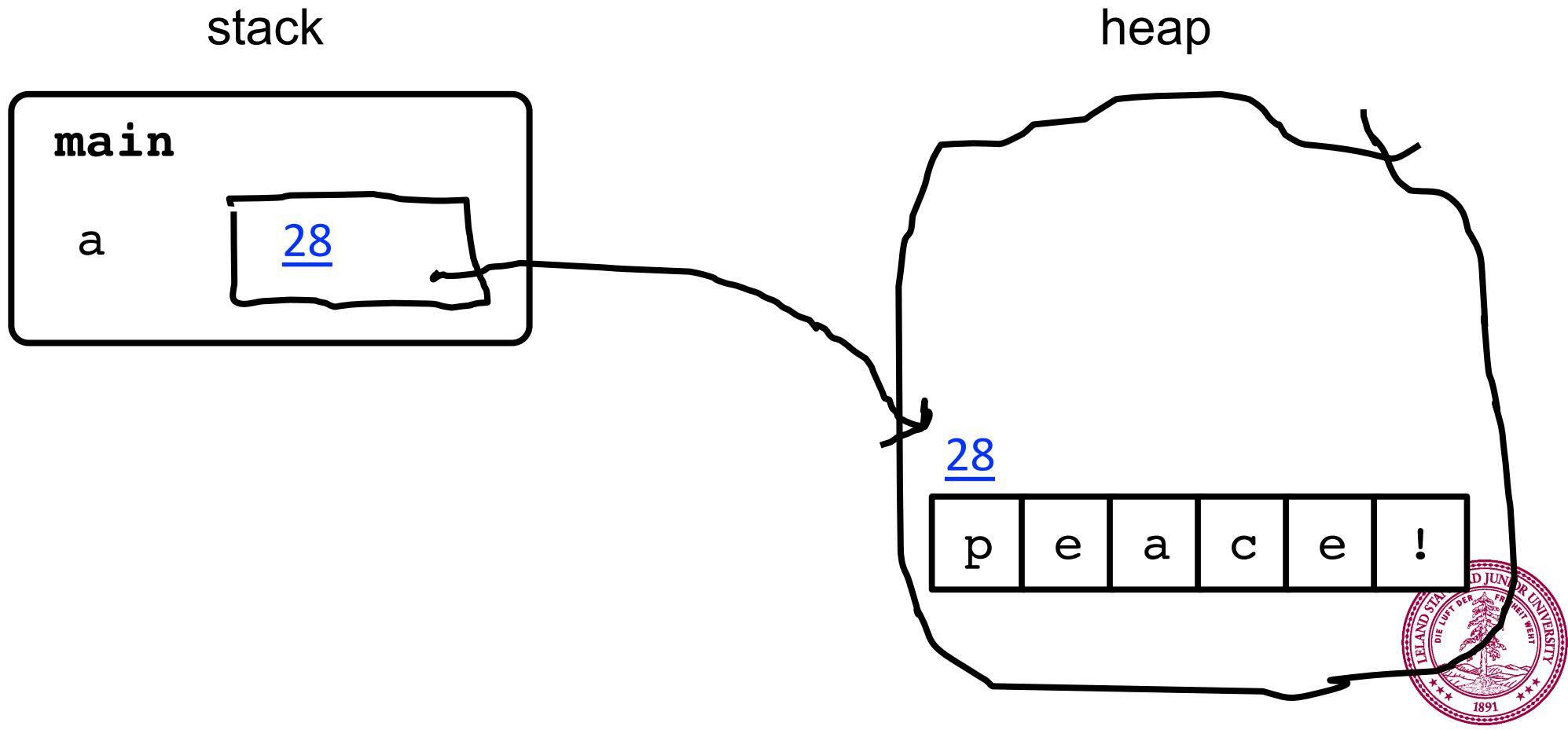
```
a = [1, 2, 3, 4, 5, 6]  
a = 'peace!'
```

stack



Aside: What is the Garbage Collector?

```
a = [1, 2, 3, 4, 5, 6]  
a = 'peace!'
```



with or without you?

6 song references later...

a lesson on garbage collection...

My Take



People bother you less
if you use the first one

```
with open('mydata.txt') as f:  
    for line in f:  
        line = line[:-1]  
        print(line)
```

```
for line in open('mydata.txt'):  
    line = line[:-1]  
    print(line)
```

But I would guess in
the future the second
one takes over...



Files + Strings = DataScience!

dataset.csv

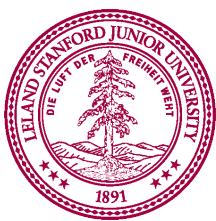
Kenya,100,50

Malaysia,50,100

...

Turkey,20,20

Spain,95,95



dataset.csv

Kenya,100,50

Malaysia,50,100

...

Turkey,20,20

Spain,95,95

```
for line in file:  
    line = line[:-1]  
    values = line.split(',')  
    print(values[1])
```



dataset.csv

Kenya,100,50

Malaysia,50,100

...

Turkey,20,20

Spain,95,95

```
for line in f:  
    line = line[:-1]  
    values = line.split(',')  
    print(values[1])
```

line

```
'Kenya,100,50\n'
```



dataset.csv

```
Kenya,100,50  
Malaysia,50,100
```

...

```
Turkey,20,20  
Spain,95,95
```

```
for line in f:  
    line = line[:-1]  
    values = line.split(',')  
    print(values[1])
```

line

```
'Kenya,100,50\n'
```



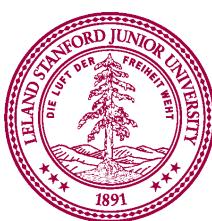
dataset.csv

```
Kenya,100,50
Malaysia,50,100
...
Turkey,20,20
Spain,95,95
```

```
for line in f:
    line = line[:-1]
    values = line.split(',')
    print(values[1])
```

line

```
'Kenya,100,50'
```



dataset.csv

```
Kenya,100,50
Malaysia,50,100
...
Turkey,20,20
Spain,95,95
```

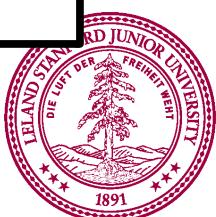
```
for line in f:
    line = line[:-1]
    values = line.split(',')
    print(values[1])
```

line

```
'Kenya,100,50'
```

values

'Kenya'	'100'	'50'
---------	-------	------



dataset.csv

```
Kenya,100,50
Malaysia,50,100
...
Turkey,20,20
Spain,95,95
```

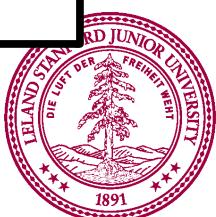
```
for line in f:
    line = line[:-1]
    values = line.split(',')
    print(values[1])
```

line

```
'Kenya,100,50'
```

values

```
'Kenya'  '100'  '50'
```



US Census Data



Thanks Keith for the cool dataset

