Compsci 101
Files, While loops, Bagels

Susan Rodger
February 10, 2022
J is for …

• **JSON**
  • Format for data transmitted across the web

• **JPEG**
  • Image format based on lossy compression

• **Jacquard Loom**
  • 1804 "automated" loom
Latanya Sweeney
PhD. Computer Science, MIT – first black woman
Over 100 publications, Fellow ACMI

“I am a computer scientist with a long history of weaving technology and policy together to remove stakeholder barriers to technology adoption. My focus is on "computational policy" and I term myself a "computer (cross) policy" scientist. I have enjoyed success at creating technology that weaves with policy to resolve real-world technology-privacy clashes.

http://latanyasweeney.org/
Identify 87% of US population using (dob,zip,gender). Prof. Government and Technology @ Harvard, instrumental in HIPAA because if de-identification work. Former CTO of the Federal Trade Comm.
One of her websites you can try: https://aboutmyinfo.org/identity

How unique am I?
Find out how much different you are among the masses.

Fill out the form below to see how unique you are, and therefore how easy it is to identify you from these values. Please note that this service is still under development.

**Date of Birth**

- **Month**
- **Day**
- **Year**

**Gender**

- Male
- Female

**ZIP Code**

ZIP code must be 5 digits long.

**Your Profile**

Results will appear here.
## Your Profile

**Gender:** Female  
**ZIP Code:** [Redacted] (pop. 46282)

<table>
<thead>
<tr>
<th>Date of Birth</th>
<th>Easily identifiable by birthdate (about 1).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth Year</td>
<td>Lots with your birth year (about 273).</td>
</tr>
<tr>
<td>Range</td>
<td>Wow! There are lots of people in the same age range as you (about 1365).</td>
</tr>
</tbody>
</table>

Five year range
Announcements

• APT-2 due tonight!
• APT-3 out, due in one week
• Assignment 2 due Tuesday, Feb. 15

• Lab 5 Friday – Do Prelab before going

• APT Quiz 1 coming … 2/24-2/27
  • APTs you take some time during this period
  • Take online
  • Two APT Quizzes – 2cd one 4/7-4/10
PFTD

- Files and Data
- While loops and Collatz sequence
- Bagel APT
Text File Processing Pattern

• See module `FileStuff.py`
  • If newline `'\n'` is read, call `.strip()`
  • If want to break line into “words”, call `.split()`

• Process the list that is returned by `.split()`
  • May need to convert strings to int or float or …

• The `for line in f:` pattern is efficient
  • Contrast list returned by `f.readlines()`
```python
def avgWord(fname):
    f = open(fname, encoding="utf-8")
    totalWords = 0
    totalLen = 0
    for line in f:
        line = line.strip()  # remove newline
        data = line.split()
        for word in data:
            totalWords = totalWords + 1
            totalLen = totalLen + len(word)
    f.close()
    return totalLen/totalWords
```
```python
def avgWord(fname):
    f = open(fname, encoding="utf-8")
    totalWords = 0
    totalLen = 0
    for line in f:
        line = line.strip()  # remove newline
        data = line.split()
        for word in data:
            totalWords = totalWords + 1
            totalLen = totalLen + len(word)
    f.close()
    return totalLen/totalWords
```
Run FileStuff

```python
if __name__ == '__main__':
    files = ['poe.txt', 'confucius.txt', 'kjv10.txt', 'oz.txt', 'species.txt']
    for f in files:
        avg = avgWord("data/"+f)
        print(f, avg)
```

Output:

```
poe.txt 4.601549053356282
confucius.txt 4.398126192817072
kjv10.txt 4.245566037162798
oz.txt 4.496446700507614
species.txt 5.036
```
Files - Summary

• Open file:   f = open(filename)
• “Process” file (2 different ways):
  • for line in f:      # get one line at a time with “\n”
  • x = f.readlines()    # x is a list of lines with “\n”
• Close file:       f.close()

• To think about when processing lines
  • Line is a string with “\n” – .strip() it
  • Maybe .split() line into list of strings (words)?
  • Convert string to int or float  - int(“376”)
When is a game of chess over?

• If you were to write a program to play chess
  • how many rounds in a game?
Another type of loop: While loop

- Repetition when you stop a loop based on a condition
- while CONDITION:
  BODY

  - As long as condition is true, keep executing loop body.
  - Must have an update in the body to get closer to condition being false
Example: while

- Playing chess

while (game not over)
make a move in the game
(game must get closer to ending)
Example: while loop – sum list

```python
lst = [4, 1, 8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)
```
Example: while loop – sum list

```
lst = [4, 1, 8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)
```

TRACE:

```
lst is [4, 1, 8]
sum 0
i 0
0<3 TRUE!
```
Example: while loop – sum list

lst = [4, 1, 8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)

TRACE:
lst  is [4, 1, 8]
sum   4
i   1

Execute body of while
Example: while loop – sum list

```
lst = [4,1,8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)
```

TRACE:

```
lst is [4, 1, 8]
sum    4
i      1
1<3    TRUE!
```
Example: while loop – sum list

```python
lst = [4, 1, 8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)
```

TRACE:

```
lst is [4, 1, 8]
sum  5
i    2
```

Execute body of while
Example: while loop – sum list

`lst = [4, 1, 8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)`

TRACE:

- lst is [4, 1, 8]
- sum = 5
- i = 2

2 < 3 TRUE!
Example: while loop – sum list

```python
lst = [4, 1, 8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)
```

```
Execute body of while
```

```
TRACE:

lst is [4, 1, 8]
sum: 13
i: 3

2/10/22 Compsci 101, Spring 2022 23
Example: while loop – sum list

lst = [4, 1, 8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)

TRACE:

lst is [4, 1, 8]
sum   13
i     3
3 < 3 FALSE! Exit while loop
Example: while loop – sum list

```
lst = [4,1,8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)
```

Output: 13
Summary: while loop

lst = [4, 1, 8]
sum = 0
i = 0

while i < len(lst):
    sum += lst[i]
i += 1

print(sum)
Summary: while loop

lst = [4,1,8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)

1) Check condition, True or false?
2) True! Execute body
3) Then check condition again. True or false?
Summary: while loop

lst = [4,1,8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)

1) Check condition, True or false?
2) False! Loop is over, go to statement following loop
History: From while to for loops

### while loop (sum list)
```
lst = [4,1,8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)
```

### for loop (sum list)
```
lst = [4,1,8]
sum = 0
for n in lst:
    sum += n
print(sum)
```
Alternative while - while True

initialize
while True:
    if something:
        break
    if something2:
        update
        update
Continue or return
Alternative while - while True

initialize

while True:
    if something:
        break
    if something2:
        update
    update

Continue or return

- while true, looks like infinite loop
- Use "if" with "break"
  - Break exits the loop
- Still need to update to get closer to exiting loop
while condition vs while True

while condition:
    body
    continue

while True:
    body
    if condition:
        break
        continue

While condition is true - must update
    - must get closer to making condition false
    - use break to exit
Compare: while - while True

```python
lst = [4,1,8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)
```

```python
lst = [4,1,8]
sum = 0
i = 0
while True:
    if i >= len(lst):
        break
    sum += lst[i]
    i += 1
print(sum)
```
Compare: while - while True

```python
lst = [4,1,8]
sum = 0
i = 0
while i < len(lst):
    sum += lst[i]
    i += 1
print(sum)
```

```python
lst = [4,1,8]
sum = 0
i = 0
while True:
    if i >= len(lst):
        break
    sum += lst[i]
    i += 1
print(sum)
```

Conditions are opposites!
WOTO-2 While loops
Now let’s see a problem that needs a while loop
Collatz Conjecture
(Hailstone)

If number is even
Divide by 2
If number is odd
multiply by 3 and add 1

Always end up at 1!
Why Solve This? In Python?

- We want to illustrate an indefinite loop
  - One of many mathematical sequences, but …

- There's an XKCD comic about it!
  - Not everyone enjoys XKCD, but …

- Mathematics is foundational in computer science, but
  - Not everyone enjoys logic/math puzzles, but …
Developing and Reasoning about While Loops

• Don't know: *how many times* loop executes
  • *a priori* knowledge, we'll know afterword

• Do know: condition that should be true after loop
  • Its negation is the expression for `BOOL_CONDITION` (loop guard)

```python
while BOOL_CONDITION:
    LOOP_BODY
    # modify variables, affect expression
```
Concrete Example: Collatz/Hailstone

• Don't know: *how many times* loop executes
  • some numbers: long sequences, others short

• Do know: condition that should be true after loop
  • It's negation is the expression for loop guard!
  • What is true after loop below finishes?

```python
while value != 1:
    loop body
    # modify value somehow
```
Collatz Code

```python
def hailstone(start, printing=False):
    steps = 0
    current = start
    while current != 1:
        if printing:
            print("{:3d}\t{:6d}".format(steps, current))
        if current % 2 == 0:
            current //= 2
        else:
            current = current * 3 + 1
        steps += 1
        if printing:
            print("{:3d}\t{:6d}".format(steps, current))
    return steps
```

What is this code doing? What gets updated? Is the loop guaranteed to stop?

What is new in this code? What does that new stuff do?
```python
def hailstone(start, printing=False):
    steps = 0
    current = start
    while current != 1:
        if printing:
            print("{:3d}\t{:6d}".format(steps, current))
        if current % 2 == 0:
            current //= 2
        else:
            current = current * 3 + 1
        steps += 1
        if printing:
            print("{:3d}\t{:6d}".format(steps, current))
    return steps
```
Collatz: New stuff

```python
def hailstone(start, printing=False):
    steps = 0
    current = start
    while current != 1:
        if printing:
            print("{:3d}\t{:6d}".format(steps, current))
        if current % 2 == 0:
            current //= 2
        else:
            current = current * 3 + 1
        steps += 1

    if printing:
        print("{:3d}\t{:6d}".format(steps, current))
    return steps
```

- **Default value, if no argument**
- **Syntax for nicer formatting**
Collatz: Guaranteed to stop?

```python
def hailstone(start, printing=False):
    steps = 0
    current = start
    while current != 1:
        if printing:
            print("{:3d}\t{:6d}\n".format(steps, current))
        if current % 2 == 0:
            current //= 2
        else:
            current = current * 3 + 1
        steps += 1

    if printing:
        print("{:3d}\t{:6d}".format(steps, current))
    return steps
```

- **current** influences the stopping condition.
- Since current is always changed, this should eventually stop.
Sample run

Output:

```python
if __name__ == '__main__':
    num = 6
    s = hailstone(num, True)
    print('num =', num, 'steps =', s)
```

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

num = 6 steps = 8
Collatz Data – Average no. of steps

• How do we gather data for numbers \( \leq 10,000 \)?
  • In general for numbers in range(low,high)?
  • Call function, store result, store 10,000 results?

• We'd like counts\([k]\) to be length of sequence for k
  • How do we allocate 10,000 list elements?
  • Like there is "hello" \(*\ 3"
  • There is \([0]\ \ast\ 10000\)
def analyze(limit):
    counts = []
    # max index into count is limit, but start at 1
    for _ in range(limit+1):
        counts.append(0)

    for n in range(1, limit+1):
        counts[n] = hailstone(n)

    avg = sum(counts)/len(counts)-1  # ignore index 0
    mx = max(counts)
    dex = counts.index(mx)
    print("average", avg)
    print("max is %d at %d" % (mx, dex))
```python
def analyze(limit):
    counts = []
    # max index into count is limit, but start at 1
    for _ in range(limit+1):
        counts.append(0)
    for n in range(1, limit+1):
        counts[n] = hailstone(n)
    avg = sum(counts)/len(counts)-1  # ignore index 0
    mx = max(counts)
    dex = counts.index(mx)
    print("average", avg)
    print("max is %d at %d" % (mx, dex))
```
counts list when limit is 8?

- Counts is of size 8+1, we ignore slot 0

- hailstone(1), get 0

- hailstone(2), get 1 step, just divide by 2

Store answer for hailstone(1) in index 1

Store answer for hailstone(2) in index 2
counts list when limit is 8

- hailstone(3), get 7 (10, 5, 16, 8, 4, 2, 1)

- hailstone(4), get 2

Store answer for hailstone(3) in index 3

Store answer for hailstone(4) in index 4
counts list when limit is 8?
• hailstone(5), get 5 (16, 8, 4, 2, 1)

• And so on…..
• Hailstone(6) is 8, hailstone(7) is 16, hailstone(8) is 3

Store answer for hailstone(5) in index 5.
WOTO-3 Collatz and While
Bagels
(Accumulation)
APT Bagels

• How figure out how many bagels needed?
• 7-steps!
APT: Bagel Counting

**Problem Statement**

You are in charge of web-based orders for your neighborhood bagel store, *The Bagel Byte*. Each evening you must total the orders to be picked up the next day. Some orders are simply for $N$ bagels, but each order of a dozen or more bagels is topped off with an extra bagel, the so-called "baker's dozen". This means, for example, that an order for 25 bagels actually requires 27 bagels to fulfill since there are two extra bagels needed for each dozen in the order. An order for 11 bagels doesn't require any extra since it's for less than a dozen.

Given a list of integers representing bagel orders determine the number of bagels needed to fulfill all the orders.

```python
filename: Bagels.py

def bagelCount(orders) :
    """
    return number of bagels needed to fulfill the orders in integer list parameter orders
    """

    # you write code here
```
Examples

1. orders = [1, 3, 5, 7]

   Returns: 16

   No order is for more than a dozen, return the total of all orders.

2. orders = [11, 22, 33, 44, 55]

   Returns: 175 since 11 + (22+1) + (33+2) + (44+3) + (55+4) = 175
Step 1 and 2

• Step 1: Solve an instance (think)
  • orders = [11, 3, 24, 17]
Step 1 and 2

• Step 1: Solve an instance (think)
  • orders = [11, 3, 24, 17]
  • $11 + 3 + (24+2) + (17+1) = 58$
  • Total: 58

• Step 2: What did we do?
  • Write down in words
WOTO-4 Step 3: Generalize
WOTO-3  Step 3: Generalize

- Go through list
  - If less than 12
    - Do nothing
  - If greater than or equal to 12
    - Add however many times 12 goes into the order
- Sum everything
Step 4: Test steps

- Go through list
  - If less than 12
    • Do nothing
  - If greater than or equal to 12
    • Add however many times 12 goes into the order

- Sum everything

  - [11, 22, 33, 44, 55]
  - 11
    • Nothing (less than 12)
  - 22
    • +1
  - 33
    • +2
  - 44
    • +3
  - 55
    • +4

  • Sum: 175
Step 5: Code

- Go through list
  - If less than 12
    - Do nothing
  - If greater than or equal to 12
    - Add however many times 12 goes into the order
- Sum everything

- for loop!
- if statement
- if’s or if...else statement?
  - floor div: //

Could we use the accumulator pattern?

Yes!
Step 5: Code

```python
def bagelCount(orders):
    """
    return number of bagels needed to fulfill
    the orders in integer list parameter orders
    """
    total = 0
    for order in orders:
        if order < 12:
            total += order
        else:
            extra = order // 12
            total += order + extra
    return total
```

- Initialize before loop
- Update inside loop
- Do something with value after loop