

Programming with the Wisdom of the Crowd

DANIEL W. BAROWY
Williams

EMERY D. BERGER
UMassAmherst

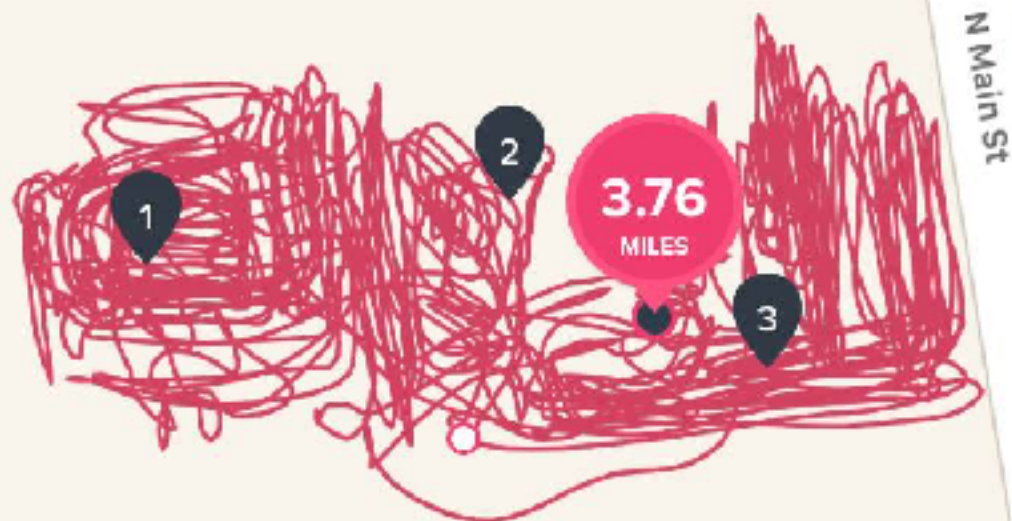
DANIEL G. GOLDSTEIN
SIDDHARTH SURI







3.76 mi 1:46:40 min 28'20" pace



Exercise

Hike

10/4/16, 8:42 AM

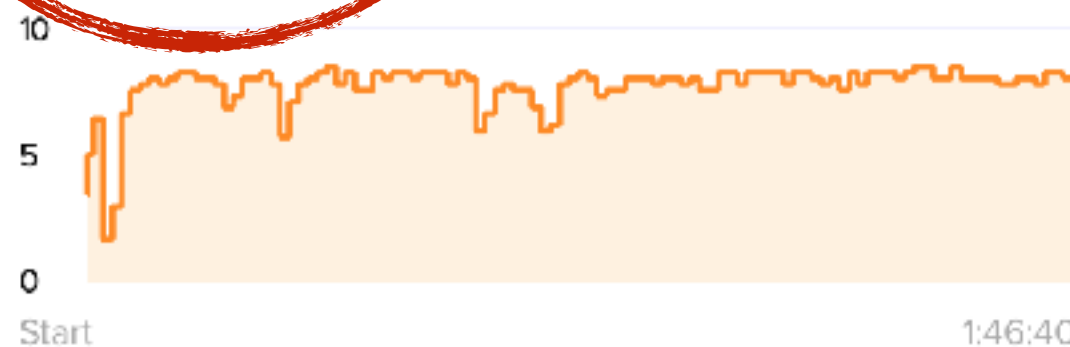


3.76 mi 1:46:40 min 28'20" pace

CALORIES BURNED

8 cals/min

835 cals



IMPACT ON YOUR DAY



+11,599
of 14,529 steps taken



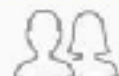
+835
of 2,717 calories burned



Dashboard



Challenges



Friends



Account



Dashboard



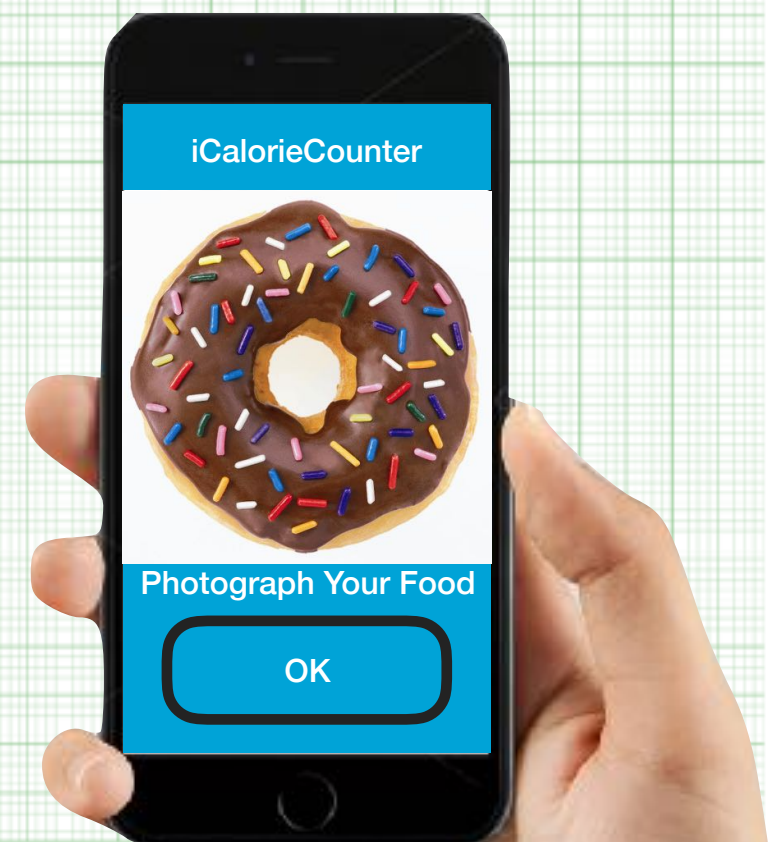
Challenges



Friends



Account





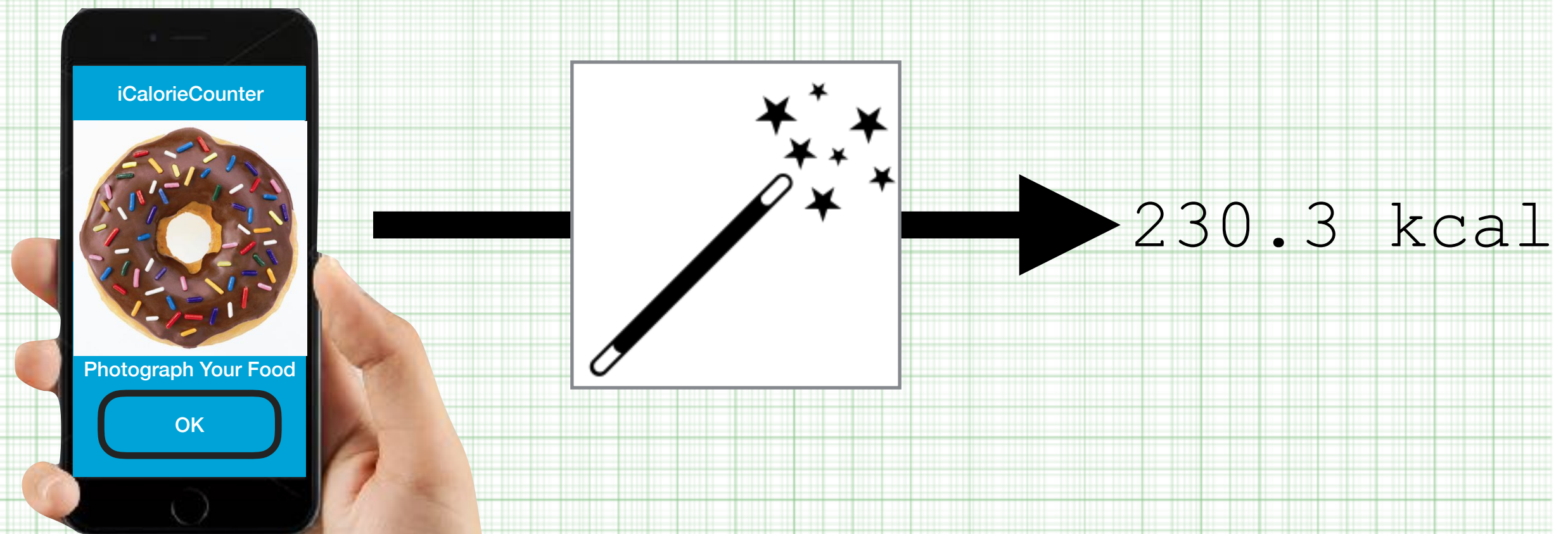
1. take a photo



→ 230.3 kcal

1. take a photo

3. return estimate



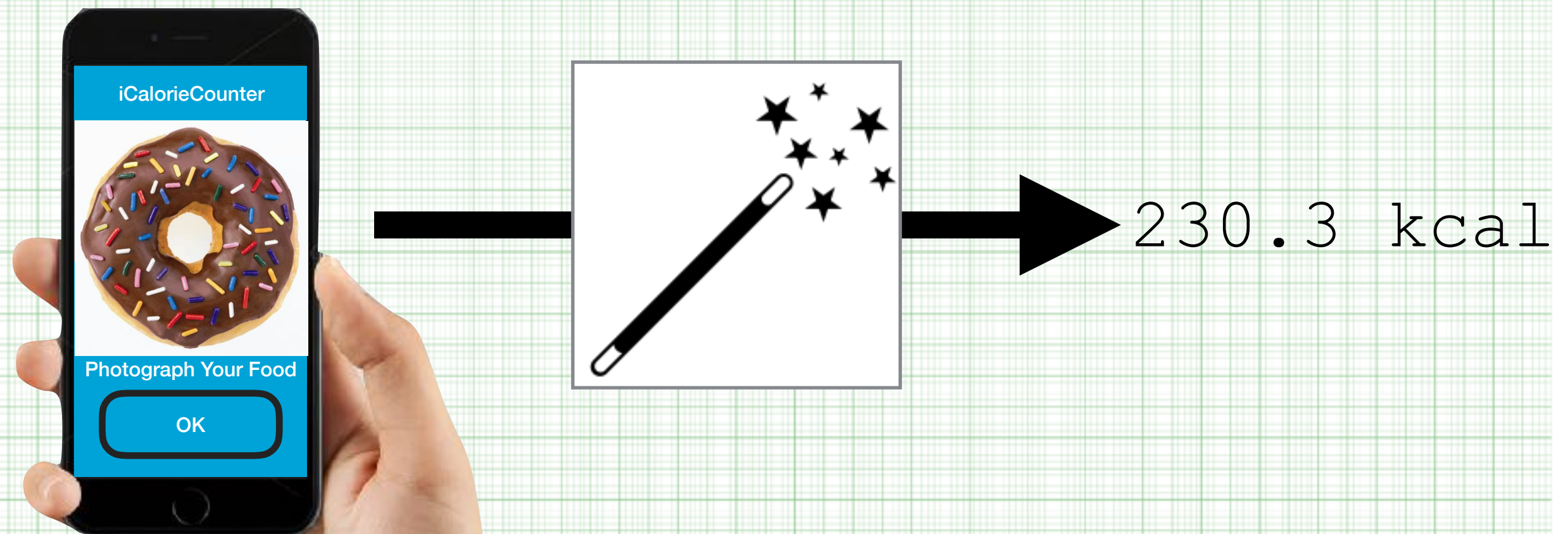
1. take a photo 2. algorithms (???) 3. return estimate



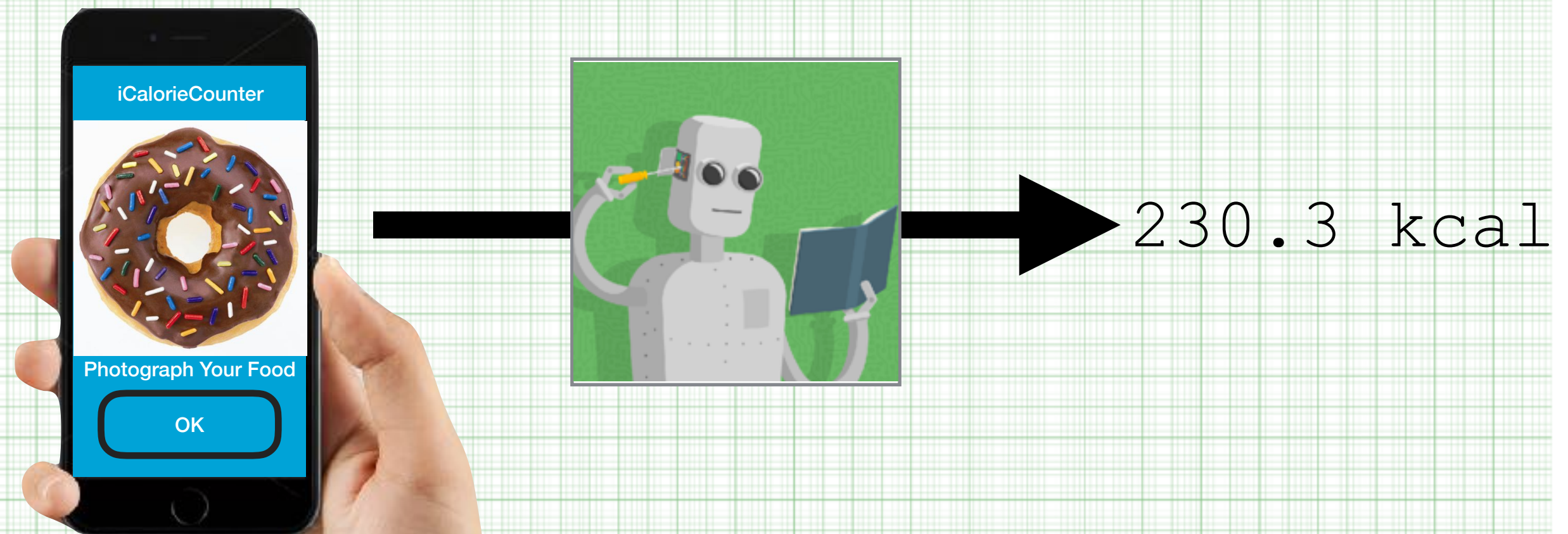
INTRODUCTION TO
ALGORITHMS
Cormen, Leiserson, Rivest

CALORIE-COUNT(a)

```
1  n = a.length
2  if n == 1
3  return a
4   $\omega_n = e^{2\pi i/n}$ 
5   $\omega = 1$ 
6   $a^{[0]} = (a_0, a_2, \dots, a_{n-2})$ 
7   $a^{[1]} = (a_1, a_3, \dots, a_{n-1})$ 
8   $y^{[0]} = \text{CALORIE-COUNT}(a^{[0]})$ 
9   $y^{[1]} = \text{CALORIE-COUNT}(a^{[1]})$ 
10 for  $k = 0$  to  $n/2 - 1$ 
11      $y_k = y_k^{[0]} + \omega y_k^{[1]}$ 
12      $y_{k+(n/2)} = y_k^{[0]} - \omega y_k^{[1]}$ 
13      $\omega = \omega \omega_n$ 
14 return  $y$ 
```

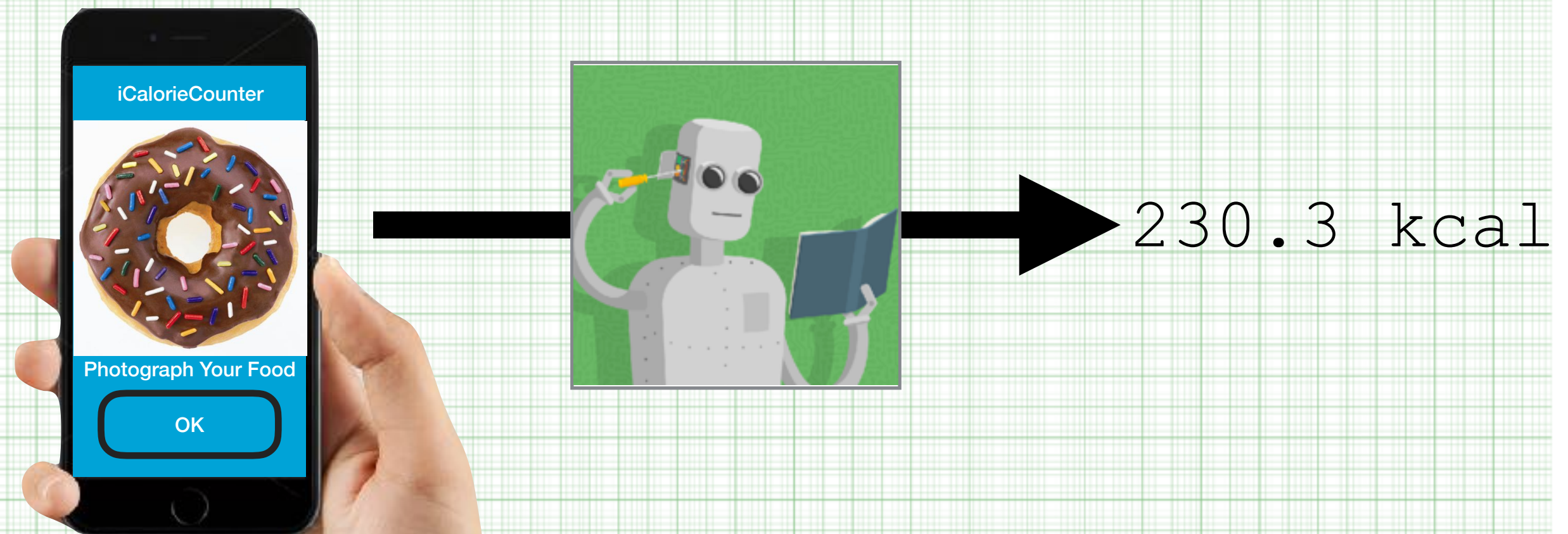



1. take a photo 2. algorithms (???) 3. return estimate

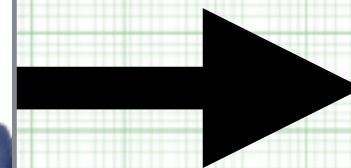
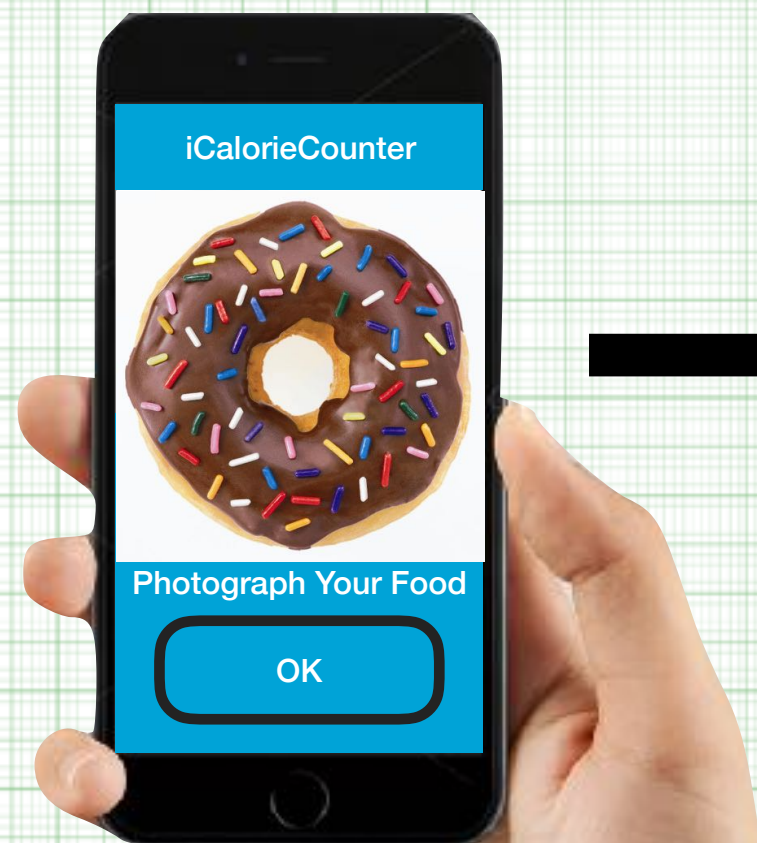


1. take a photo 2. machine learning 3. return estimate





1. take a photo 2. machine learning 3. return estimate



230.3 kcal

1. take a photo

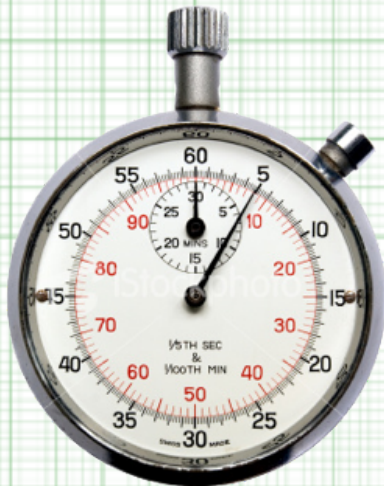
2. humans

3. return estimate

VoxPL

A programming language for **estimates** that lets you
utilize  workers as if they are just ordinary

functions in  **Scala** and  **Java** .



(like cloud functions!)

Automatically handles scheduling, payment, and quality.

Discover, preview and complete HITs on the new Worker website. [Try it out Today!](#)

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.

422,206 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



[Find HITs Now](#)

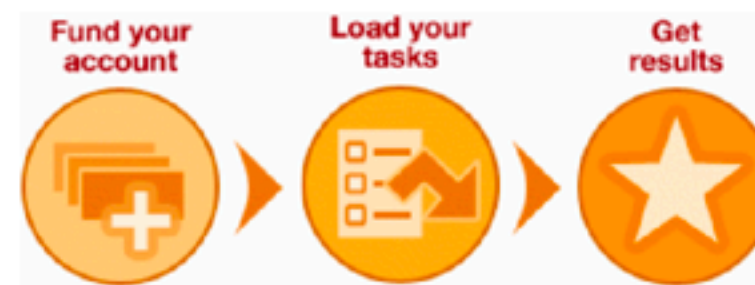
or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Get Started.](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results



[Get Started](#)

Discover, preview and complete HITs on the new Worker website. Try it out Today!

Your Account

HITs

Qualifications

425,483 HITs
available now

All HITs | HITs Available To You | HITs Assigned To You

Find HITs containing

that pay at least \$ 0.00

☐ for which you are qualified

☐ require Master Qualification



All HITs

1-10 of 1638 Results

Sort by: HITs Available (most first)

Show all details | Hide all details

1 2 3 4 5 > Next

Items per Page: 10

Type text you see in images

[View a HIT in this group](#)

Requester: Amazon Requester Inc. - Rekognition Team

HIT Expiration Date: Nov 12, 2017 (1 week 6 days)

Reward: \$0.03

Time Allotted: 2 hours

Face Landmark Annotations

[View a HIT in this group](#)

Requester: Amazon Requester Inc. - Rekognition Team

HIT Expiration Date: Nov 13, 2017 (2 weeks)

Reward: \$0.00

Time Allotted: 3 hours

Komodo Product Classification Service Audit : Marketplace 4

[View a HIT in this group](#)

Requester: A3N4U4MIJ5ZNJS

HIT Expiration Date: Nov 26, 2017 (3 weeks 6 days)

Reward: \$0.00

Time Allotted: 4 weeks 2 days

Komodo Product Classification Service Audit : Marketplace 3

[View a HIT in this group](#)

Requester: A3N4U4MIJ5ZNJS

HIT Expiration Date: Nov 26, 2017 (3 weeks 6 days)

Reward: \$0.00

Time Allotted: 4 weeks 2 days

Komodo Product Classification Service Audit : Marketplace 5

[View a HIT in this group](#)

Requester: A3N4U4MIJ5ZNJS

HIT Expiration Date: Nov 26, 2017 (3 weeks 6 days)

Reward: \$0.00

Time Allotted: 4 weeks 2 days

Transcribe up to 35 Seconds of Media to Text - Earn up to \$0.17 per HIT!!

[View a HIT in this group](#)

Requester: Crowdsurf Support

HIT Expiration Date: Oct 28, 2018 (51 weeks 6 days)

Reward: \$0.05

Your Account

HITs

Qualifications

390,009 HITs
available now

All HITs | HITs Available To You | HITs Assigned To You

Find HITs containing

that pay at least \$ 0.00

☐ for which you are qualified☐ require Master Qualification

Timer: 00:00:18 of 2 hours

Finished with this HIT?

Let someone else do it?

Submit HIT

Return HIT

☐ Automatically accept the next HITTotal Earned: \$0.12
Total HITs Submitted: 2

Type text you see in images

Requester: Amazon Requester Inc. - Recognition Team

Qualifications Required: Location is not one of: US-AL, US-AK, US-HI, US-IA, US-ID, US-LA, US-ME, US-MO, US-MS, US-NE, US-NM, US-OK, US-RI, US-SD, US-UT, US-VT, US-WY

Reward: \$0.03 per HIT

HITs Available: 26751

Duration: 2 hours

becomes	something	ADVENTURE	jewels
and	LOVE	grow	goes
GUINNESS	who	som	-William



becomes

Type the text above into the line below.

Respect capitalization, spacing and punctuation, or your work will be rejected.

0/19 words

Skip (Tab)



Save (Enter)

Finished with this HIT?

Submit HIT

Let someone else do it?

Return HIT

Your Account

HITs

Qualifications

390,009 HITs
available now

All HITs | HITs Available To You | HITs Assigned To You

Find HITs containing

that pay at least \$ 0.00

☐ for which you are qualified☐ require Master Qualification

Timer: 00:00:18 of 2 hours

Finished with this HIT?

Let someone else do it?

Submit HIT

Return HIT

☐ Automatically accept the next HITTotal Earned: \$0.12
Total HITs Submitted: 2

Type text you see in images

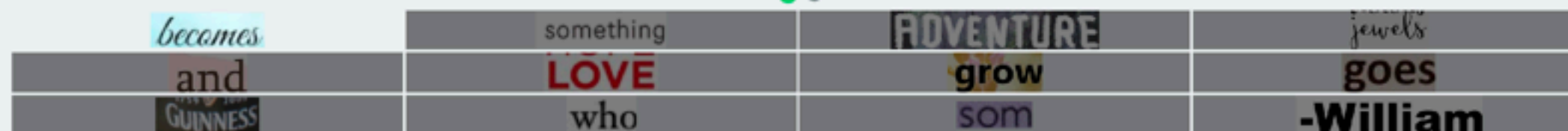
Requester: Amazon Requester Inc. - Recognition Team

Qualifications Required: Location is not one of: US-AL, US-AK, US-HI, US-IA, US-ID, US-LA, US-ME, US-MO, US-MS, US-NE, US-NM, US-OK, US-RI, US-SD, US-UT, US-VT, US-WY

Reward: \$0.03 per HIT

HITs Available: 26751

Duration: 2 hours



Type the text above into the line below.

Respect capitalization, spacing and punctuation, or your work will be rejected.

becomes

0/19 words

Skip (Tab)



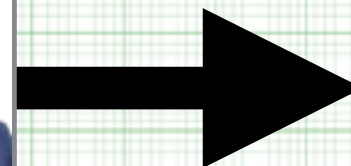
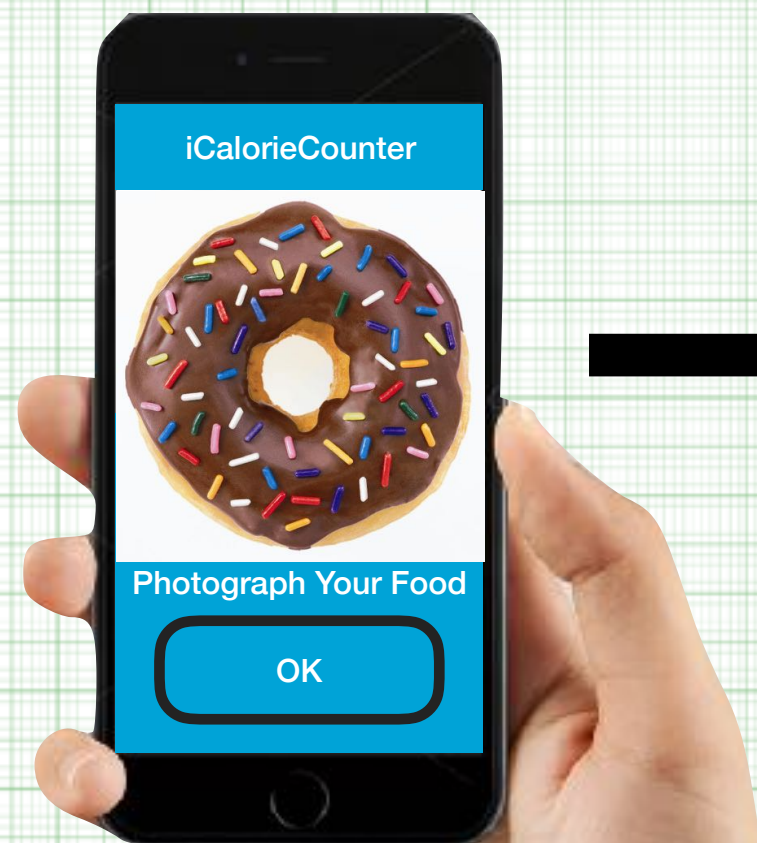
Save (Enter)

Finished with this HIT?

Submit HIT

Let someone else do it?

Return HIT

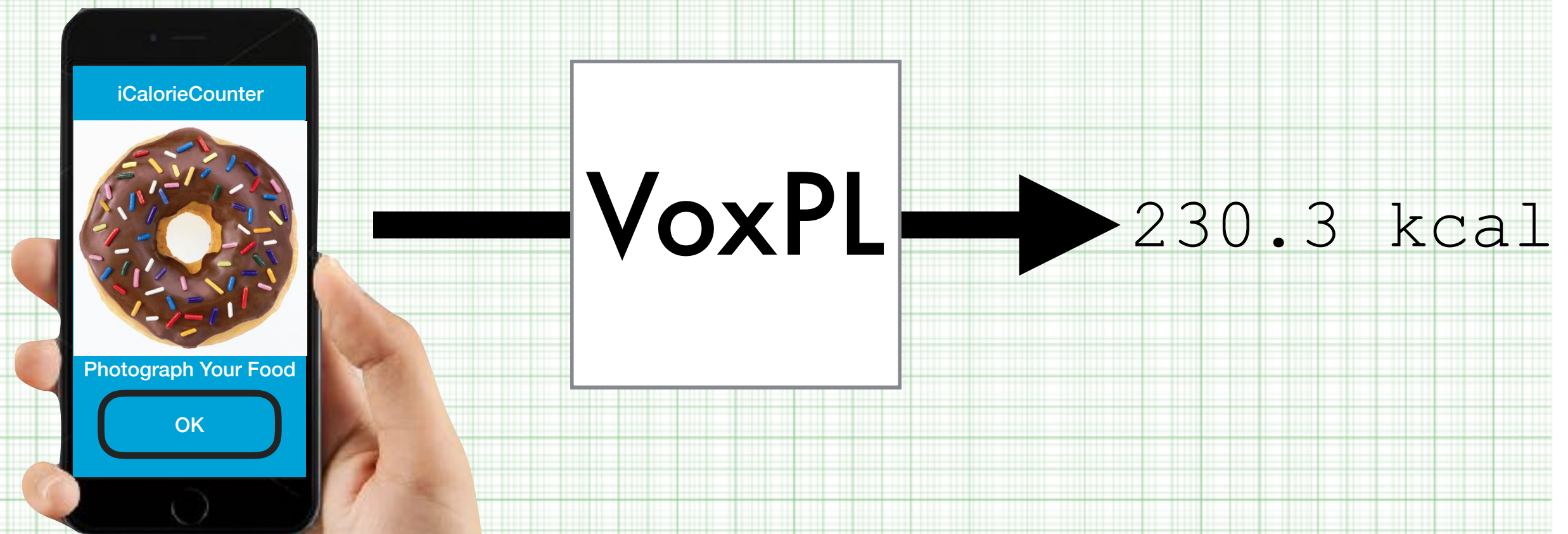


230.3 kcal

1. take a photo

2. crowdsourcing

3. return estimate



1. take a photo
2. crowdsourcing
3. return estimate

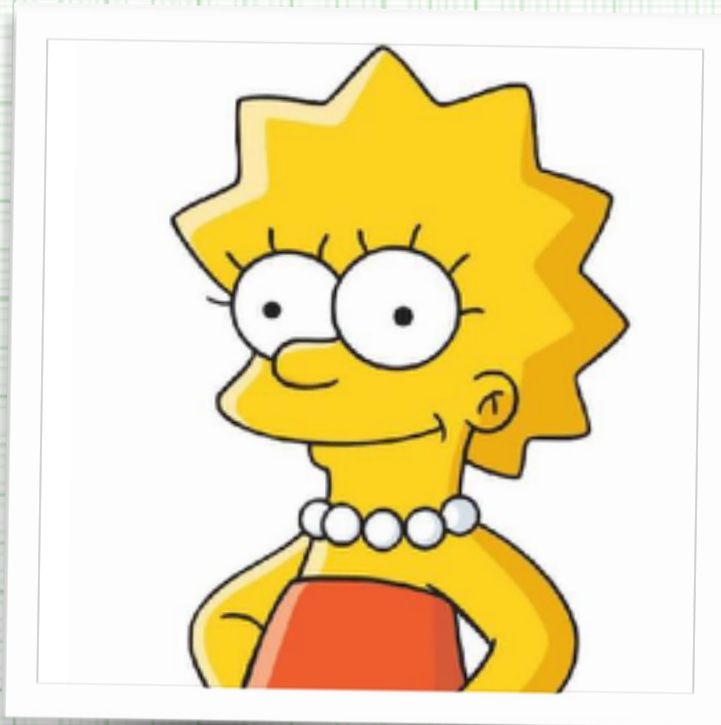
Why VoxPL?

Challenge: work quality



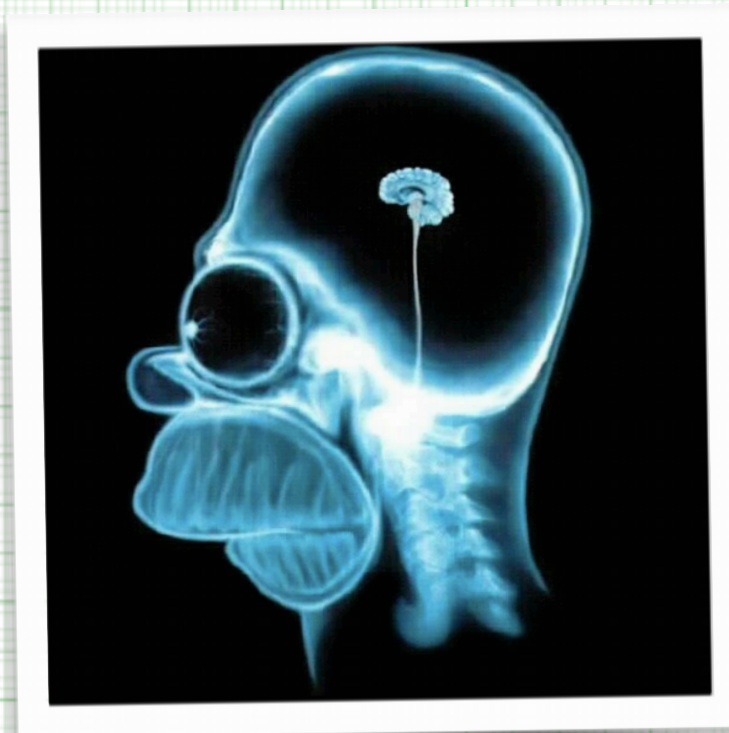
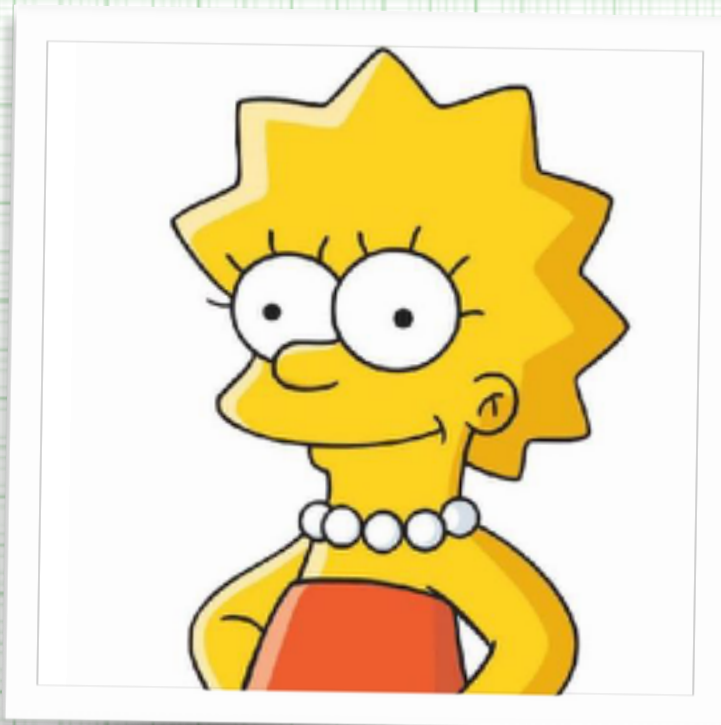
Why VoxPL?

Challenge: work quality



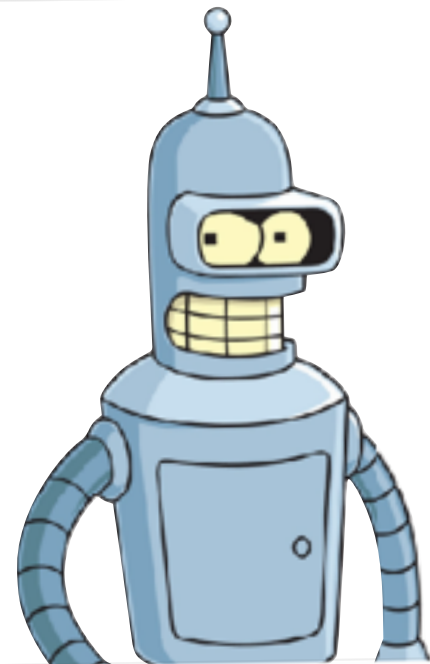
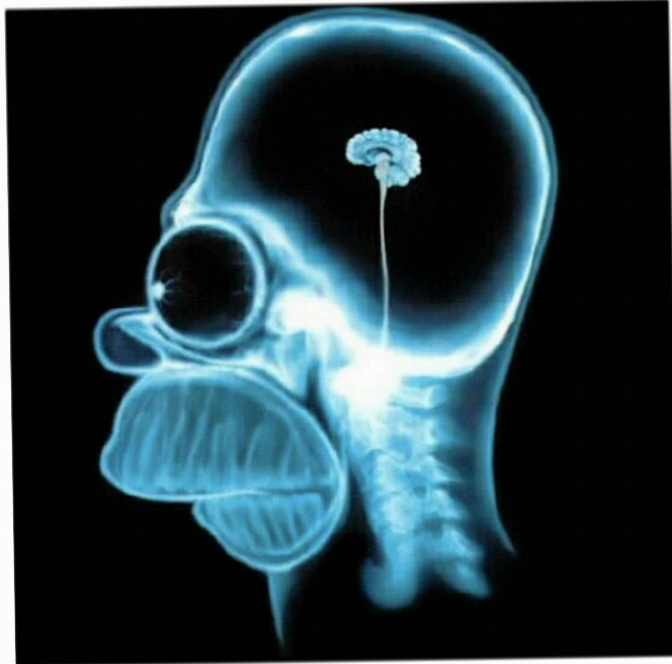
Why VoxPL?

Challenge: work quality



Why VoxPL?

Challenge: work quality



Why VoxPL?

Challenge: work quality



i.e., do workers agree?



Is this a giraffe?



no



yes



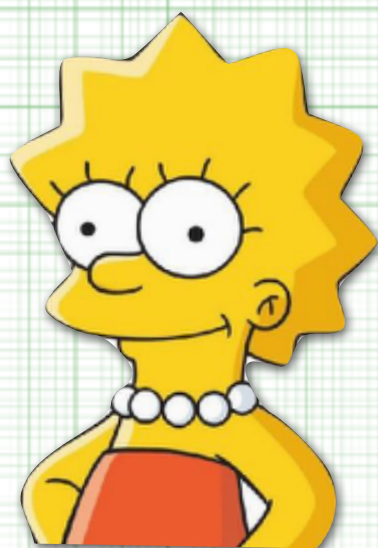
Is this a giraffe?



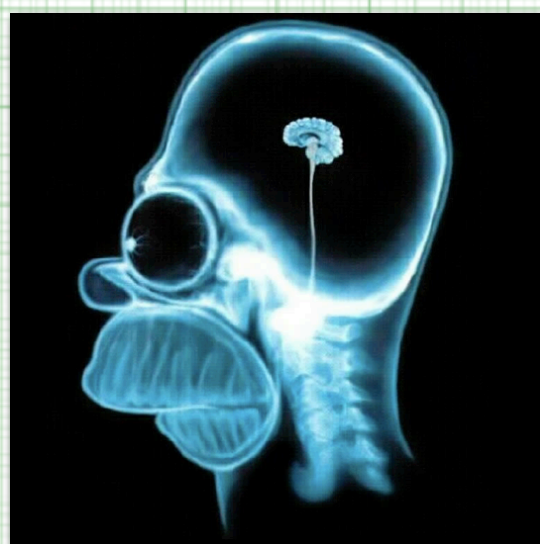
no



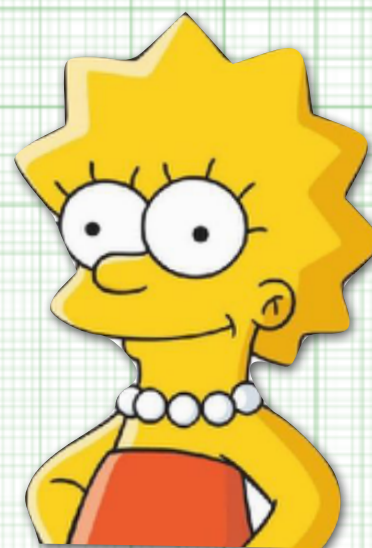
yes



yes



no



yes



Is this a giraffe?

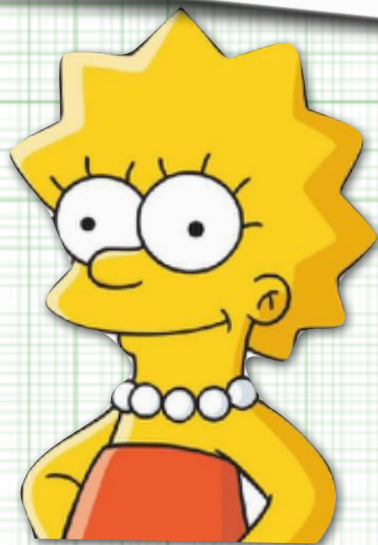


no

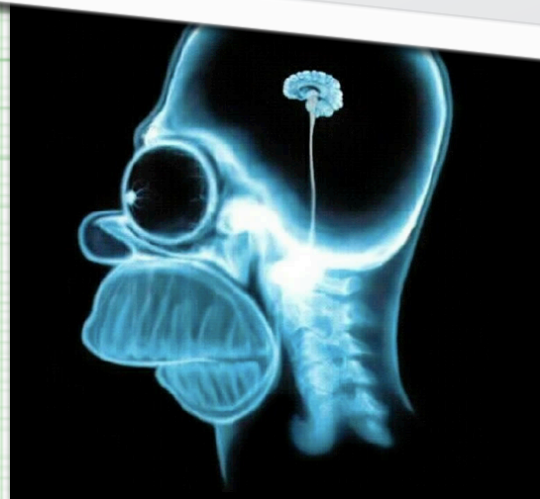


yes

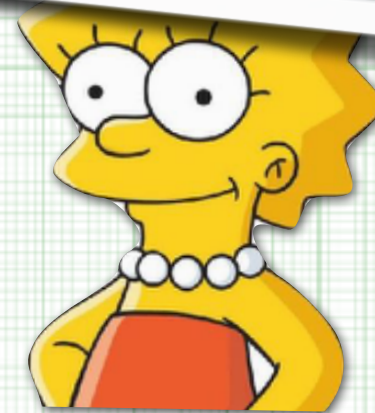
Take the majority opinion



yes



no



yes

Labeling tasks

Is this a giraffe?

☒ Yes

☐ No

Which ones don't belong?

- ☐ Oscar the Grouch
- ☐ Cookie Monster
- ☐ Kermit the Frog
- ☒ Spongebob Squarepants
- ☐ The Count

What does this plate say?



XXXXXX

≈ answer choices drawn from a finite,
low-cardinality set.

Labeling tasks

Is this a giraffe?



Yes

Which ones don't belong?

☐

Oscar the Grouch

☐

Cookie Monster

What does this plate say?



"majority opinion" is inadequate
(AutoMan: A Platform for Integrating Digital and Human Computation,
Barowy, Curtsinger, Berger, McGregor; OOPSLA '12)

≈ answer choices drawn from a finite,
low-cardinality set.



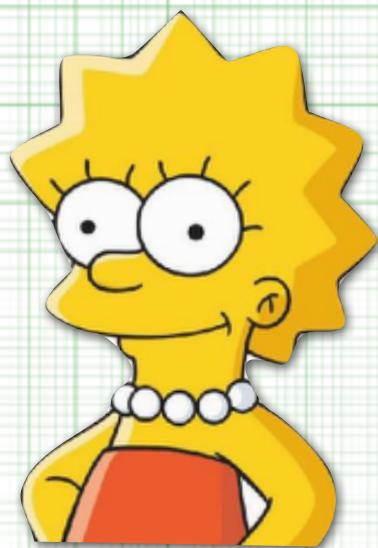
Is this a giraffe?



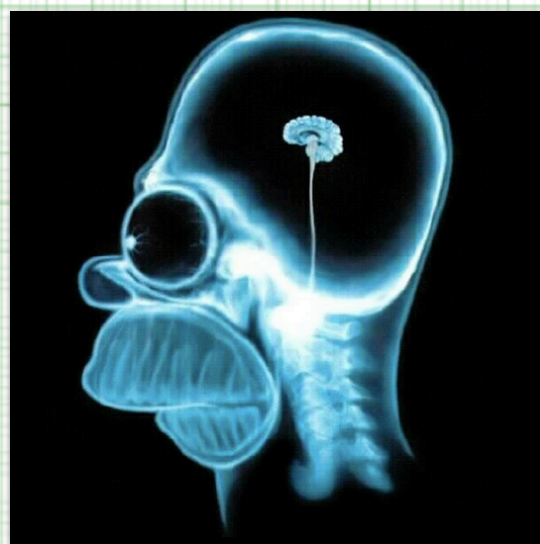
no



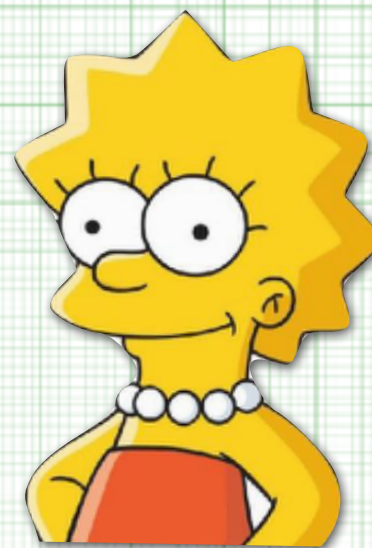
yes



yes



no



yes



Is this a giraffe?



no



yes





Is this a giraffe?



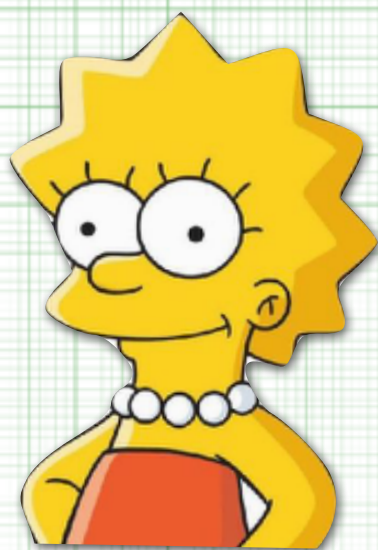
no

$P(\text{majority}) = 1 !!!$

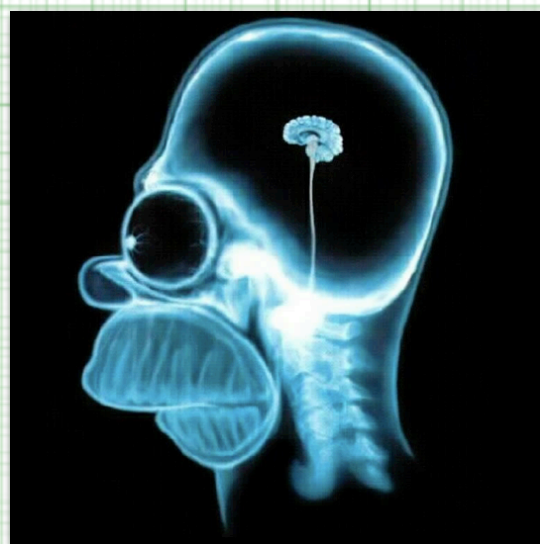




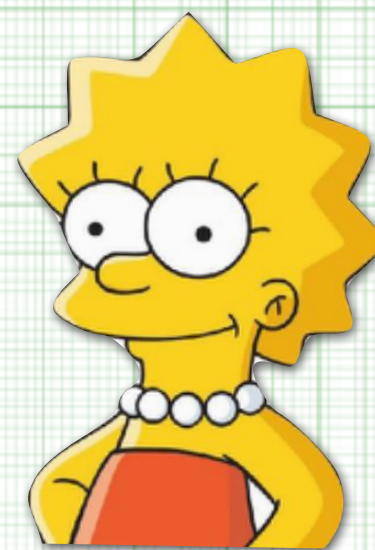
How many calories?



331



10

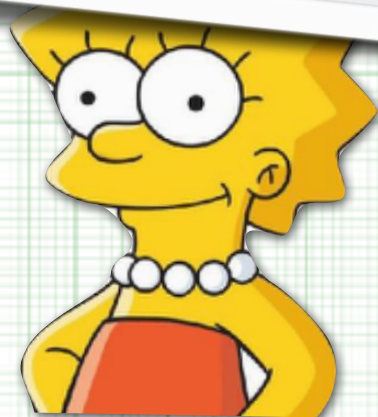


352

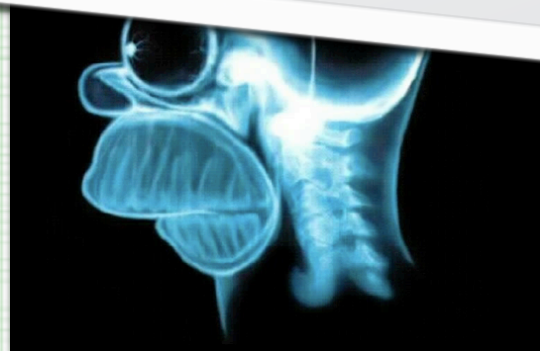


How many calories?

"majority opinion" is meaningless



331



10



352

Estimation tasks

How many calories?



Where is this person's nose?



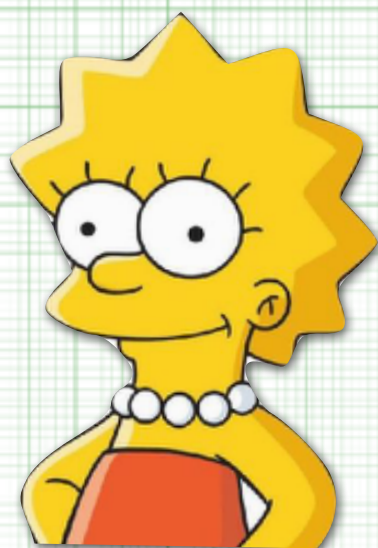
How hot is it in this photo?



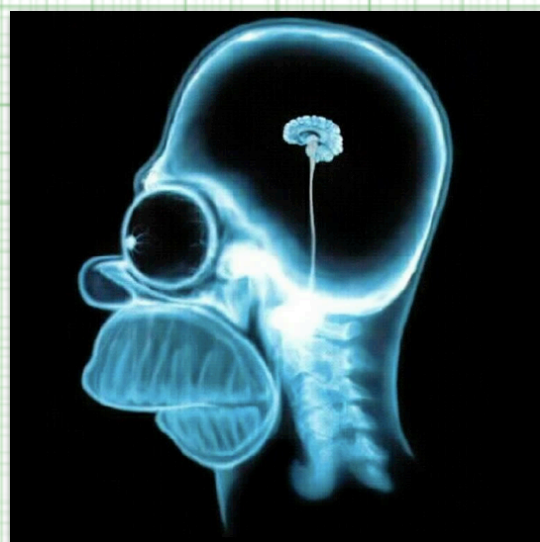
$\approx$ answer choices drawn from an infinite,
(or practically infinite) high-cardinality set.



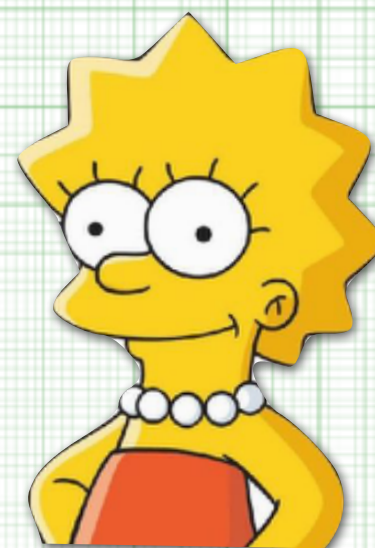
How many calories?
mean = 231



331



10



352



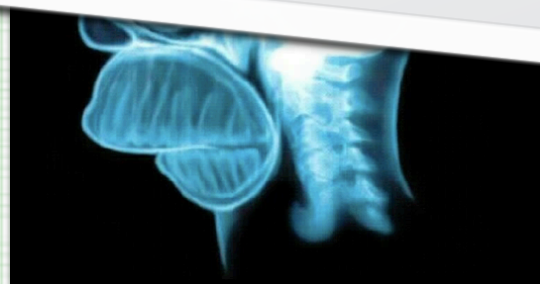
How many calories?

mean = 231

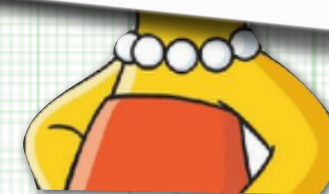
Challenge: outliers
One large outlier can skew the mean!



331



10



352



How many calories?

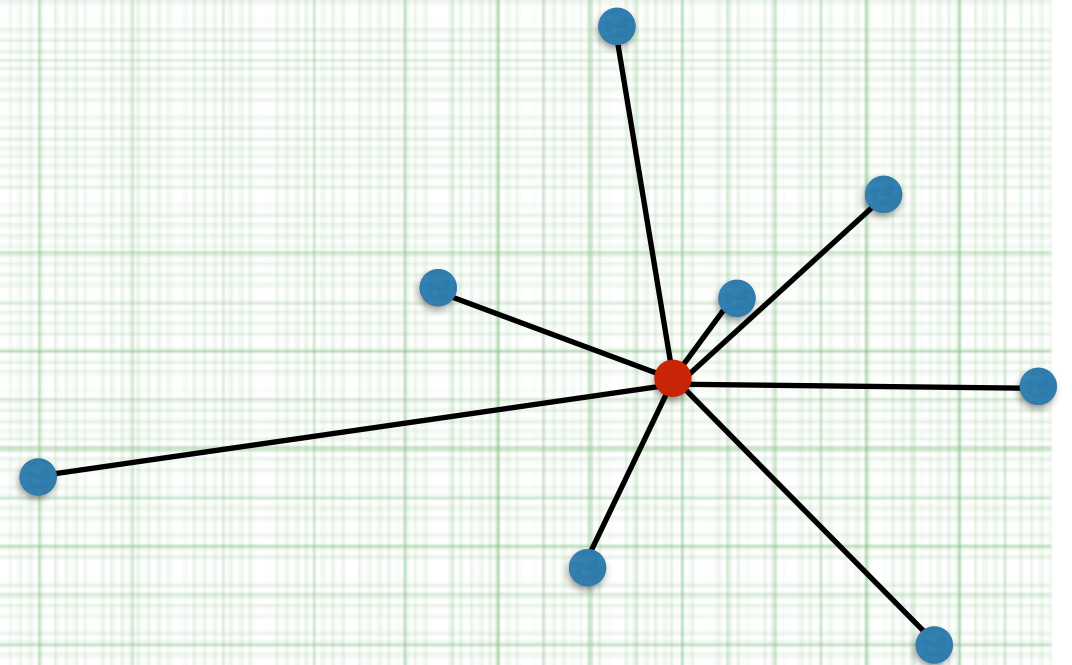
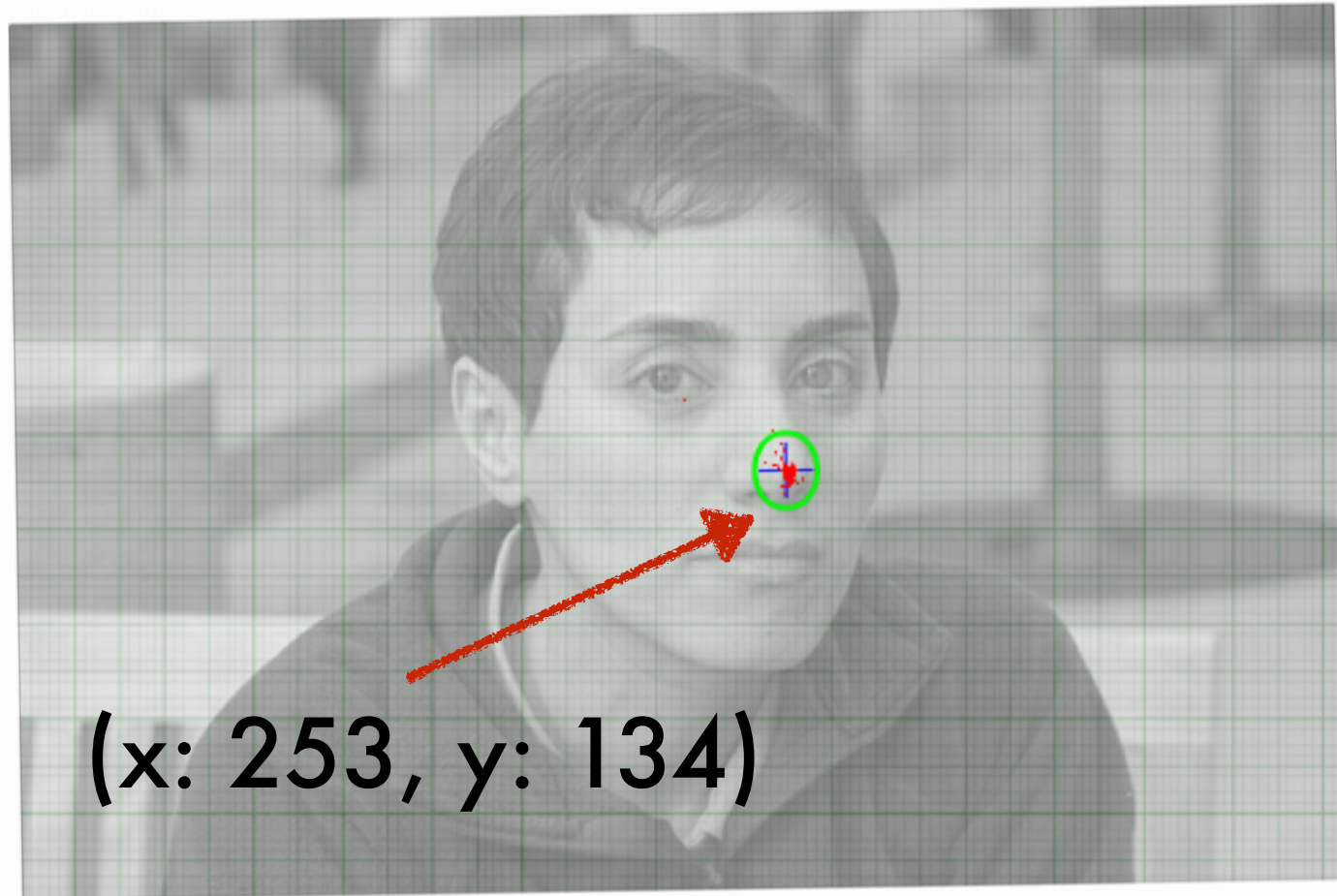
mean = 231

median = 331

Challenge: outliers
One large outlier can skew the mean!

**median: must corrupt more than 1/2
of sample to skew estimate.**

VoxPL default estimator: L_1 median



L_1 median: point that minimizes distance to all other points



**Challenge:
what is a “good”
estimate?**



“Good” is domain-specific.



**A few extra sprinkles don't matter
for calorie counting!**



**What would it take
to trust that the median value of
244 is a good estimate?**

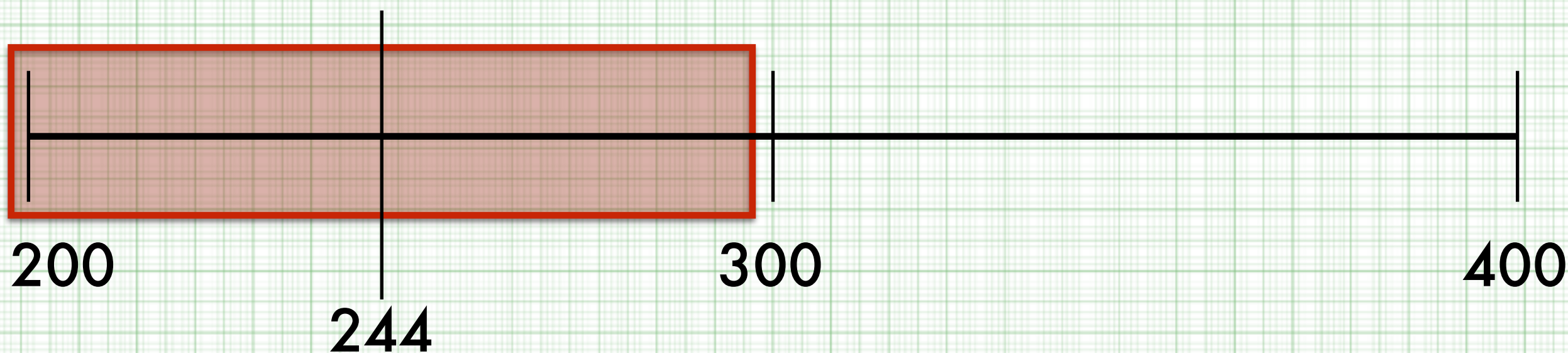


**What would it take
to trust that the median value of
244 is a good estimate?**

**95 out of 100 times we ask the crowd, the
estimate is between 194 and 294 kcal.**



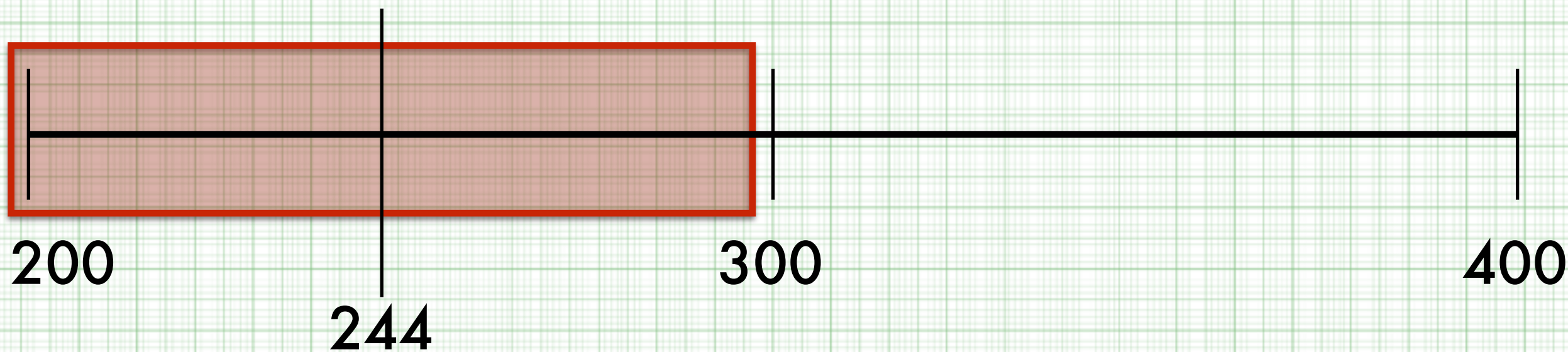
What would it take to trust that the median value of 244 is a good estimate?



95 out of 100 times we ask the crowd, the estimate is between 194 and 294 kcal.



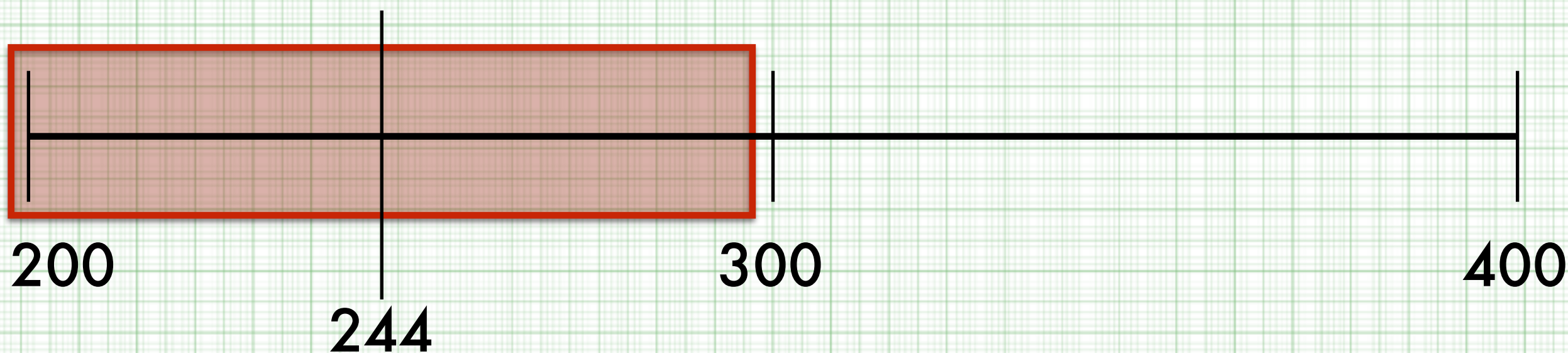
What would it take
to trust that the median value of
244 is a good estimate?



I.e., a confidence interval.



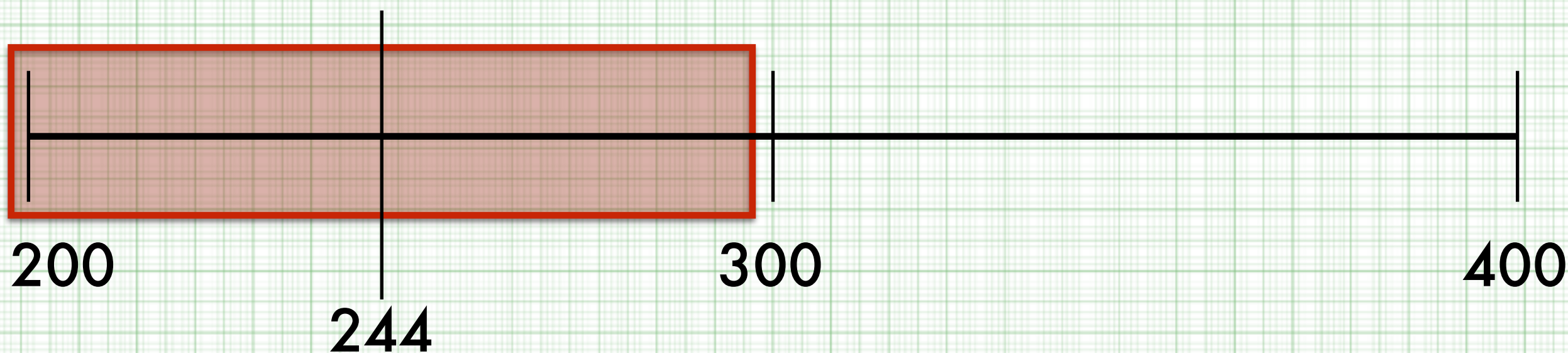
What would it take
to trust that the median value of
244 is a good estimate?



The set of estimates that are indistinguishable
due to sampling error.



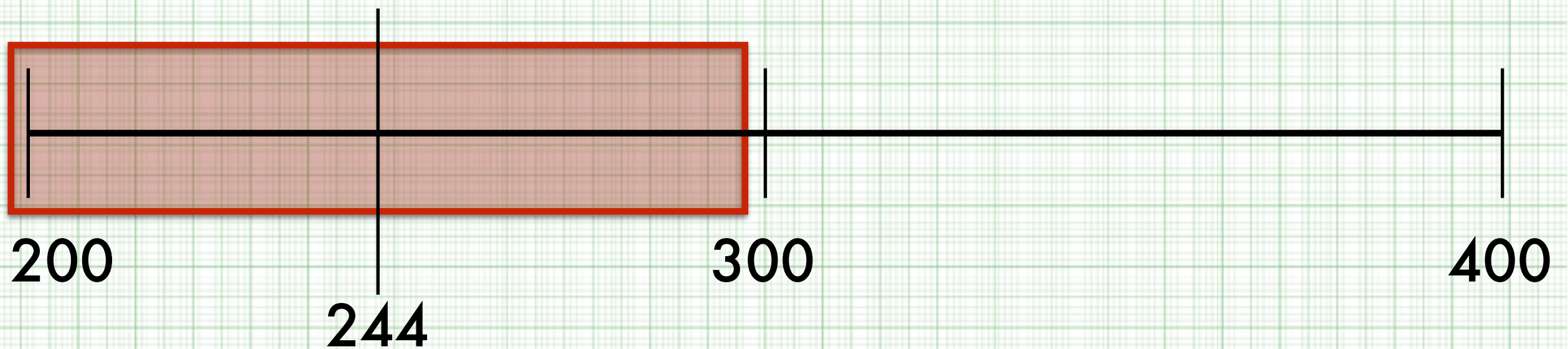
What would it take
to trust that the median value of
244 is a good estimate?



The set of opinions that are not likely to belong
to the Homers or the Benders.



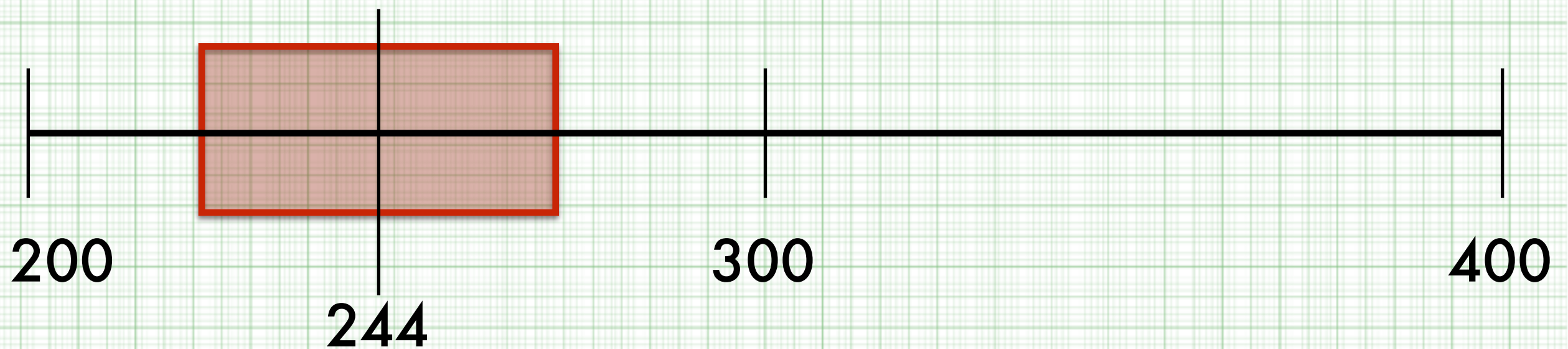
What would it take
to trust that the median value of
244 is a good estimate?



“Donut contains 244 ± 50 kcal.”



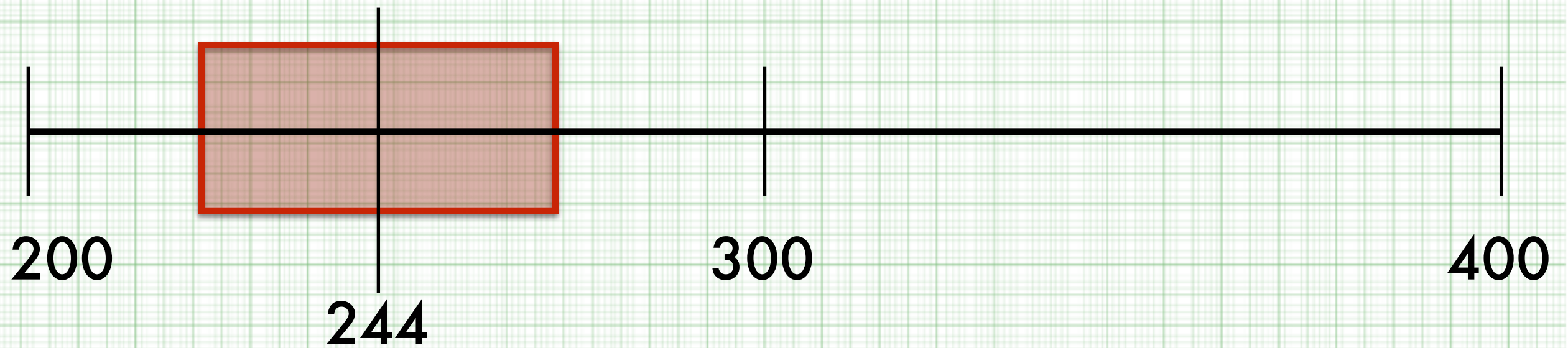
What would it take
to trust that the median value of
244 is a good estimate?



“Donut contains 244 ± 25 kcal.”



How to get tighter estimates?

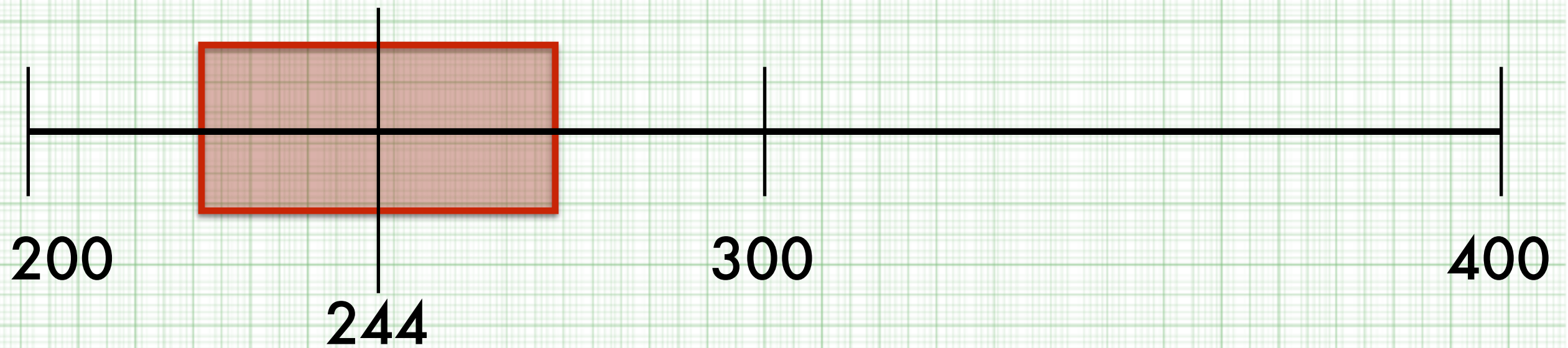


"Donut contains 244 ± 25 kcal."



How to get tighter estimates?

Ask more people.

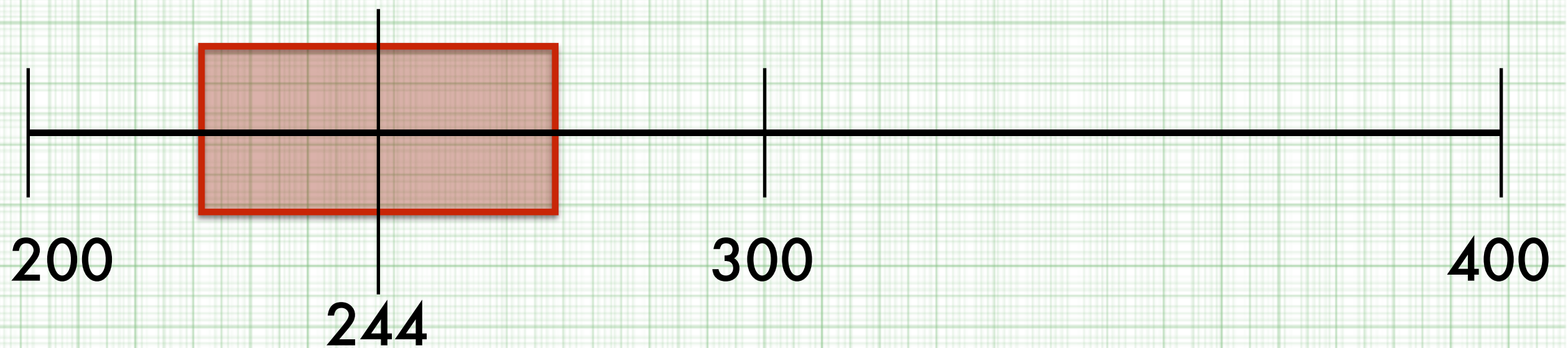


“Donut contains 244 ± 25 kcal.”



How to get tighter estimates?

Error decreases as sample size increases.

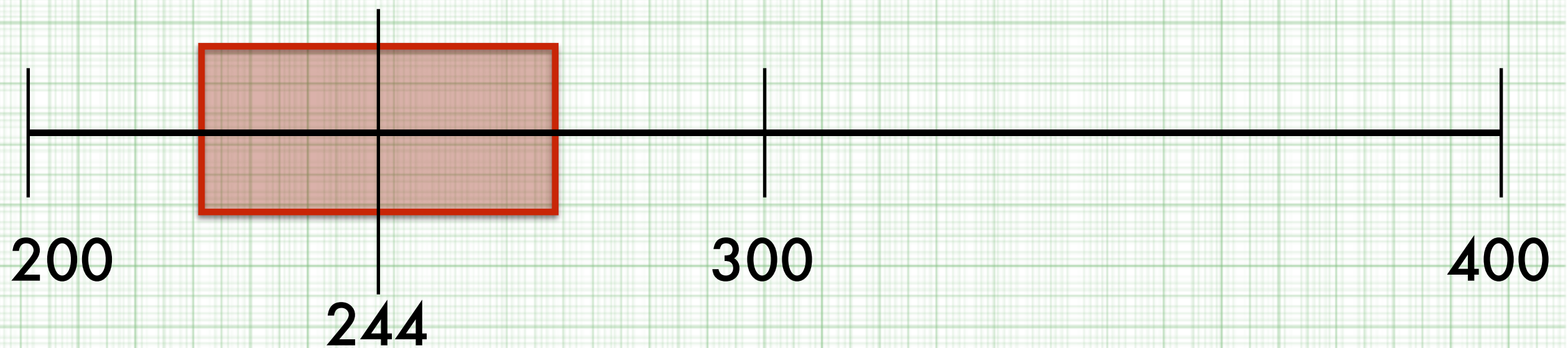


“Donut contains 244 ± 25 kcal.”



How to get tighter estimates?

Error decreases as sample size increases.



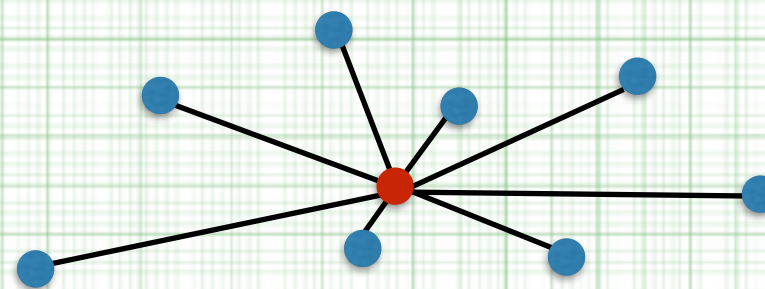
“Donut contains 244 ± 25 kcal.”
 ± 25 calories is a great estimate!



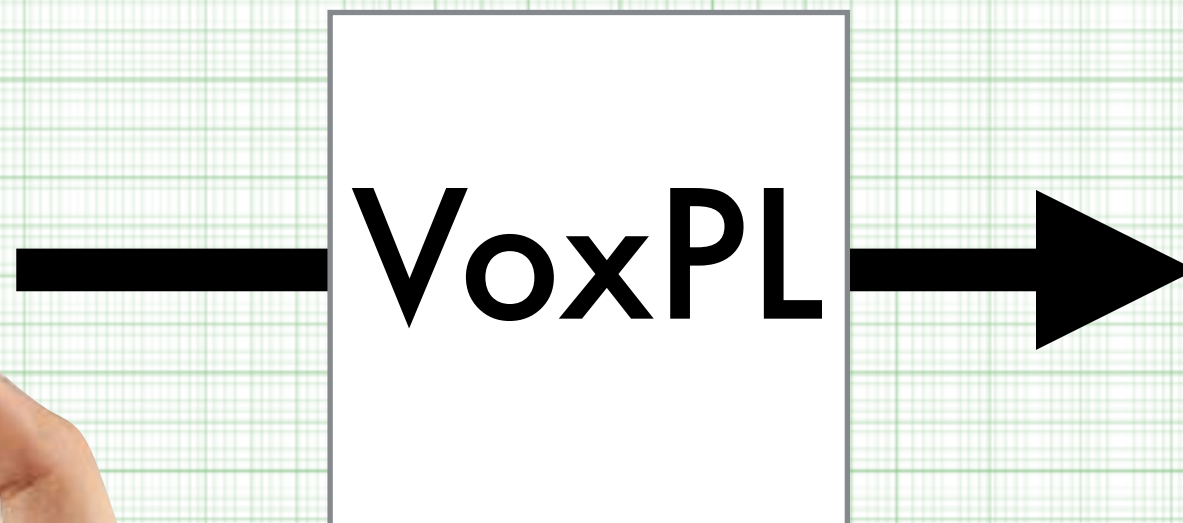
\$5.00



± 25 kcal



L1 median

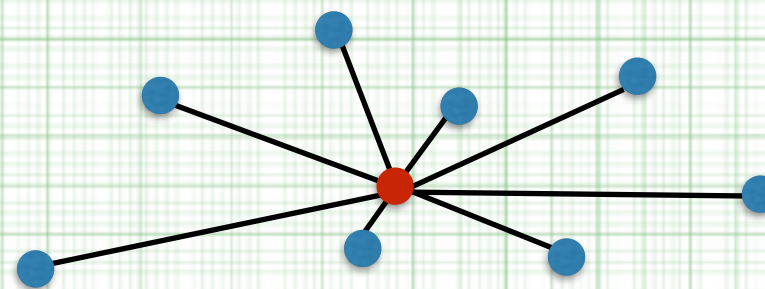




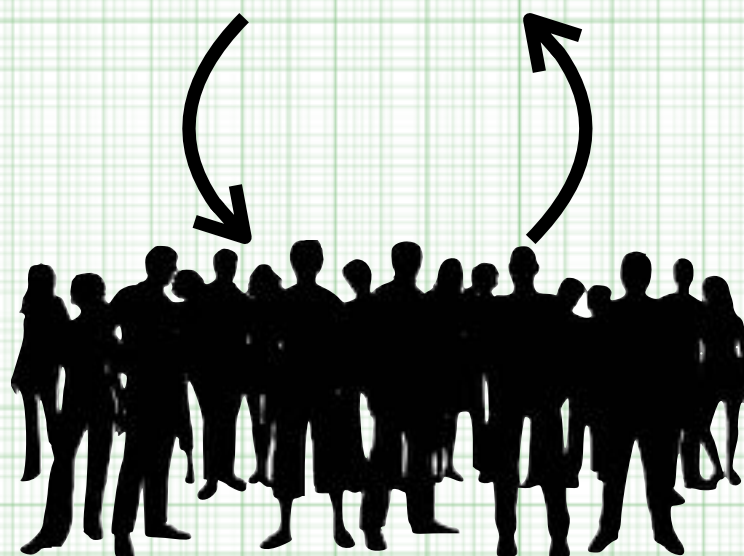
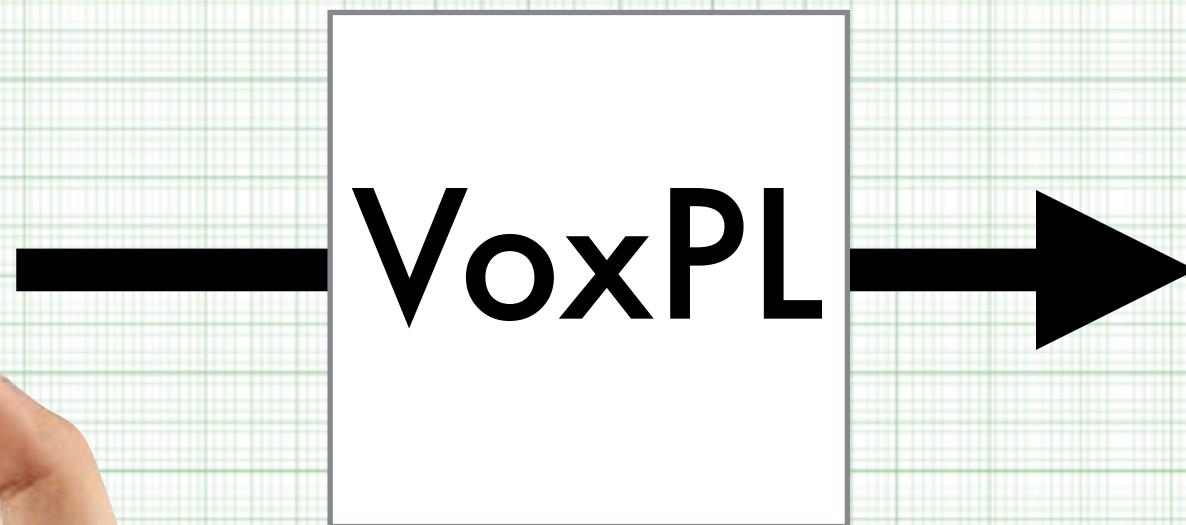
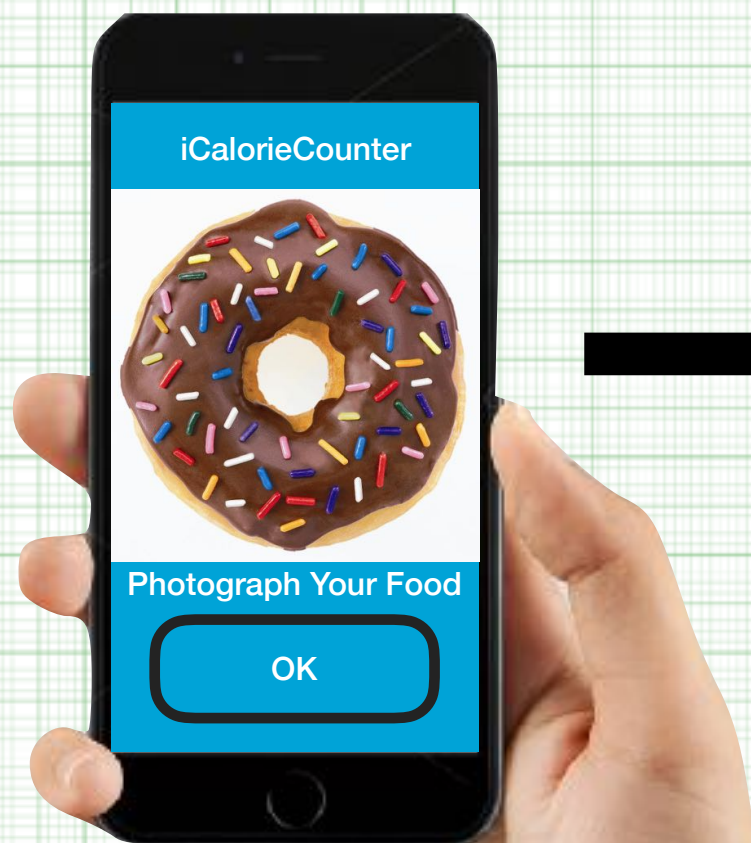
\$5.00



± 25 kcal



L1 median



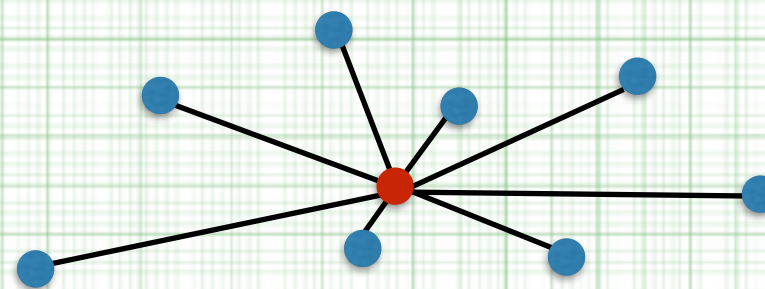
230.3 ± 111



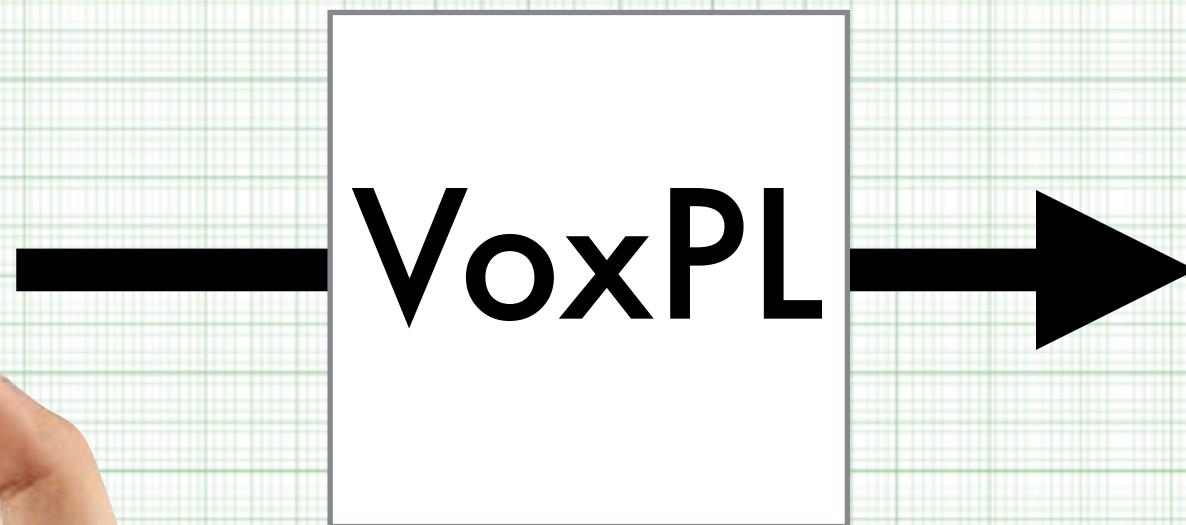
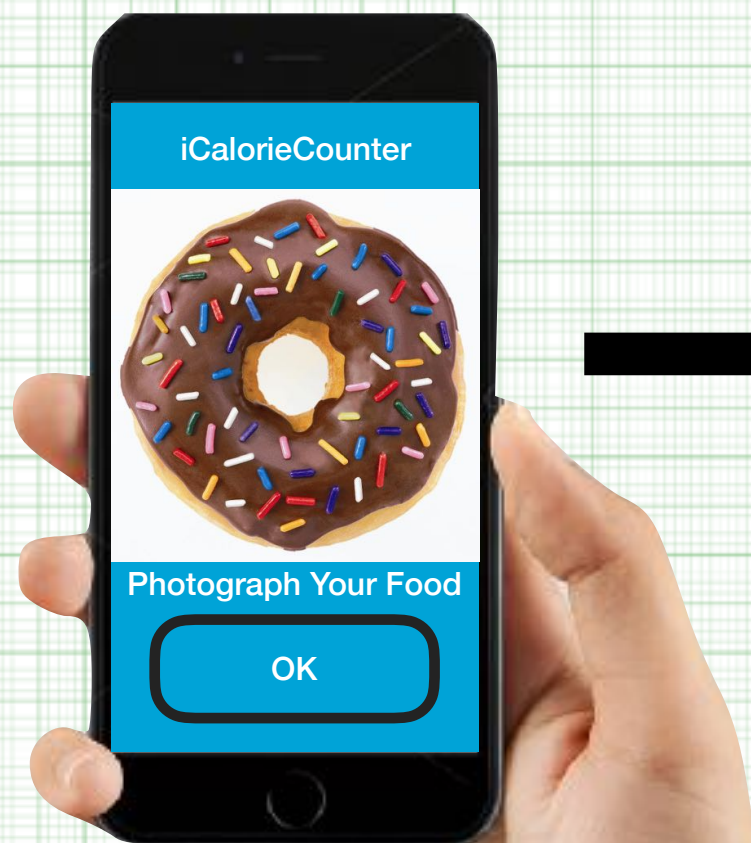
\$5.00



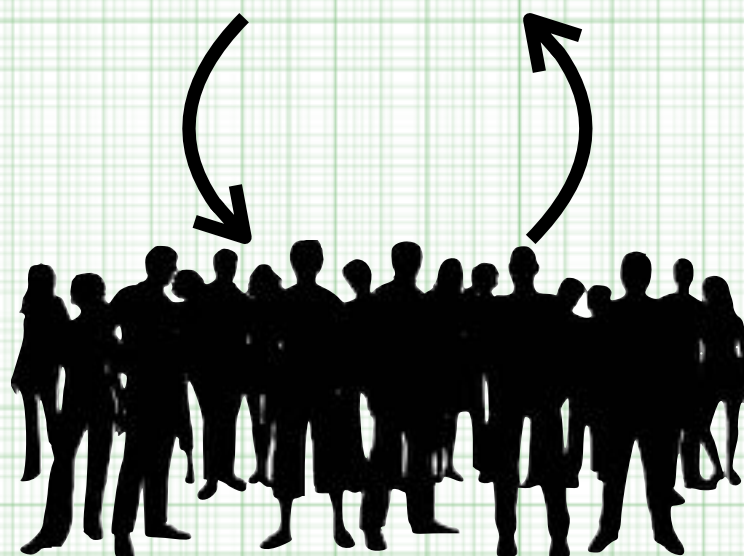
± 25 kcal



L1 median



230.3 ± 111
 $> \pm 25$

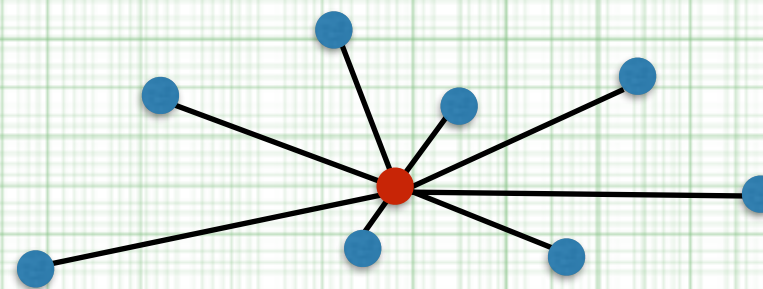




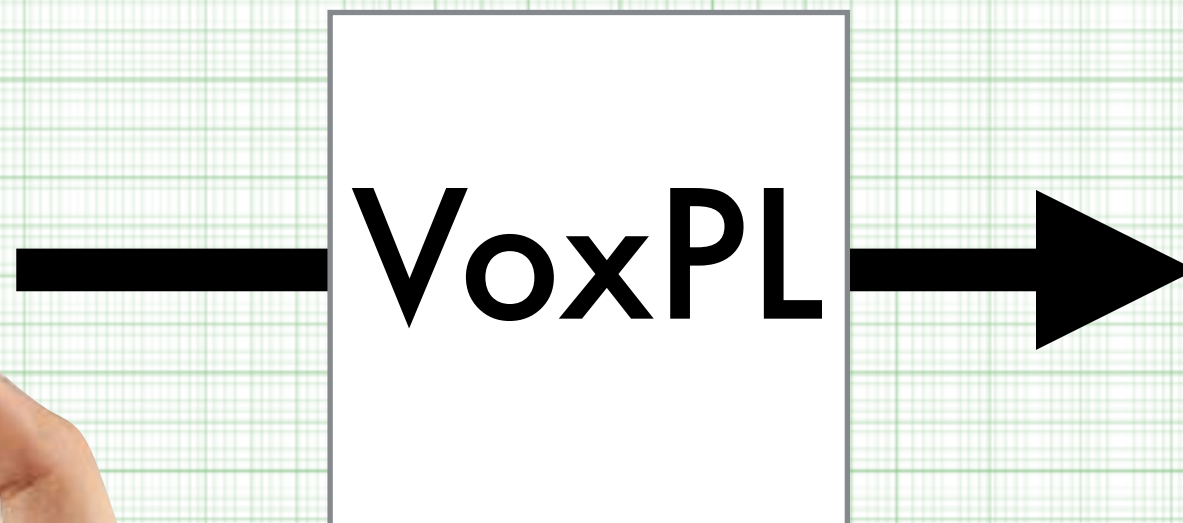
\$5.00



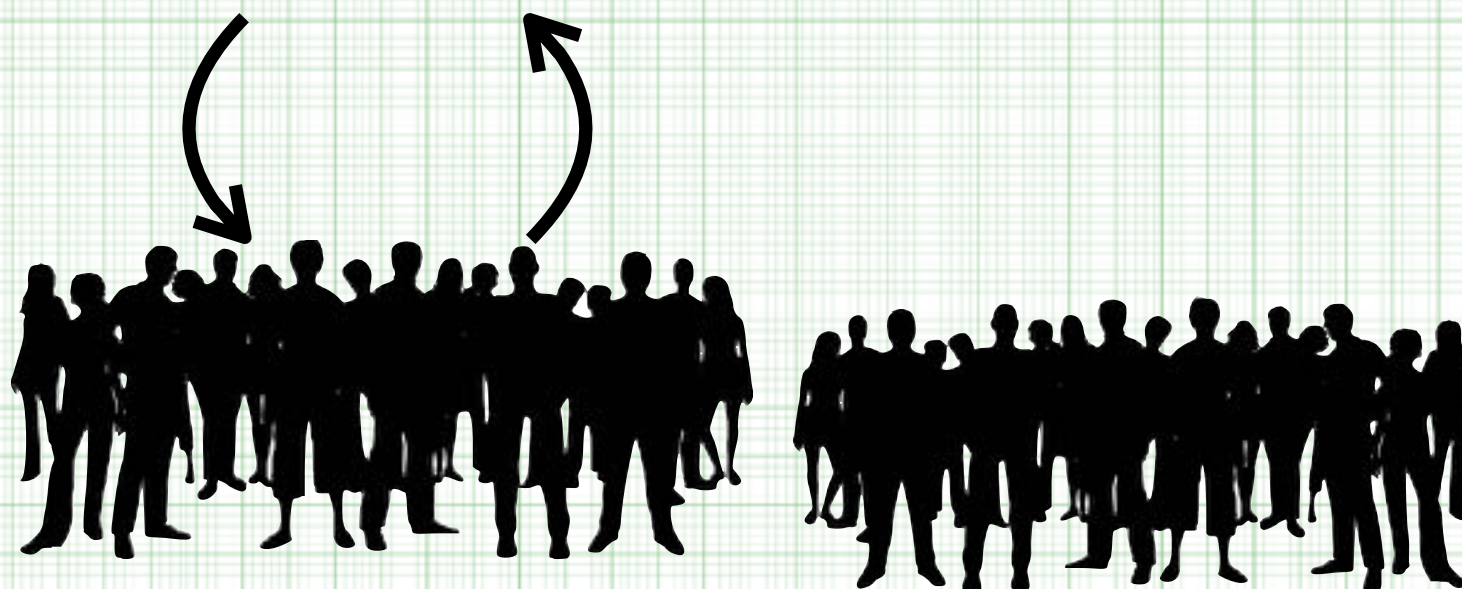
± 25 kcal



L1 median



230.3 ± 49
 $> \pm 25$

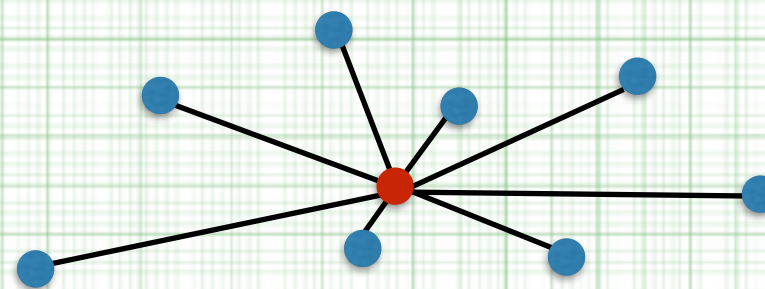




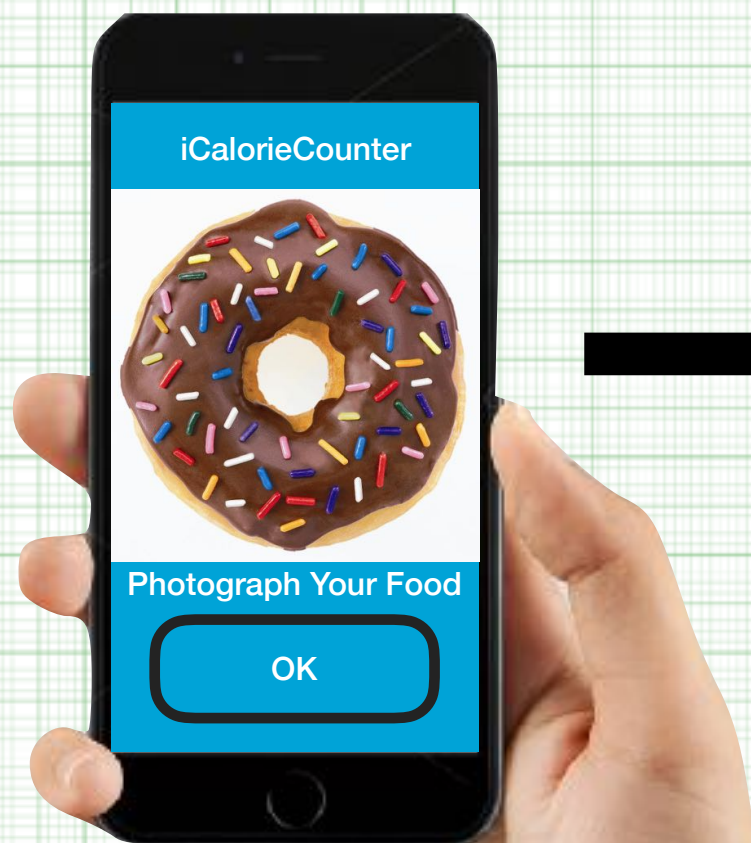
\$5.00



± 25 kcal



L1 median



VoxPL

230.3 ± 22

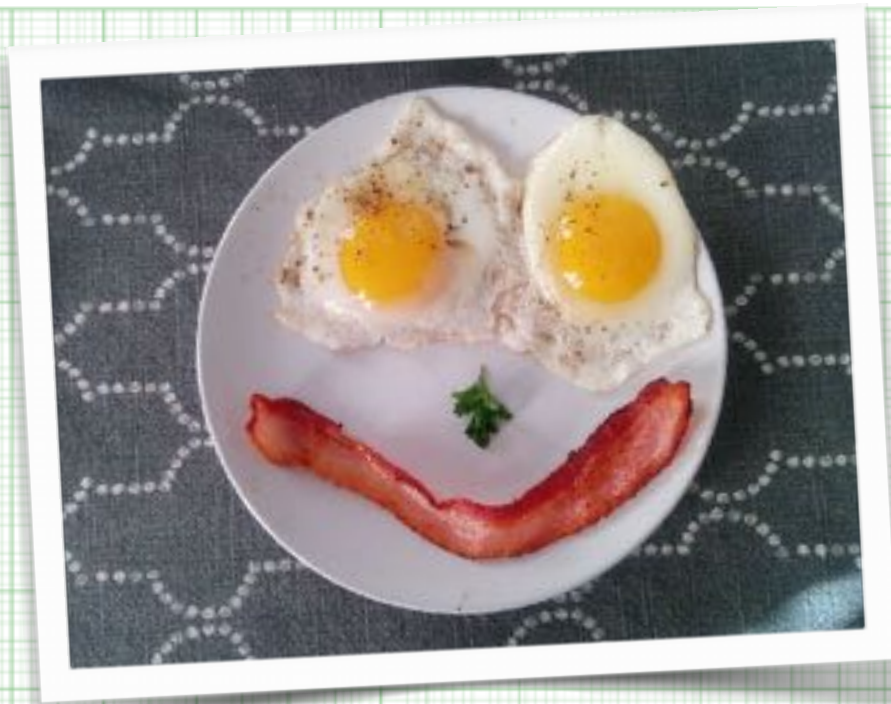
good!



VoxPL

```
def numCalories(url: String) =  
  estimate (  
    budget = 5.00,  
    confidenceInterval = SymmetricCI(25),  
    text = "How many calories are  
           in the food pictured?",  
    imageUrl = url  
  )
```





VoxPL also lets you compose estimates

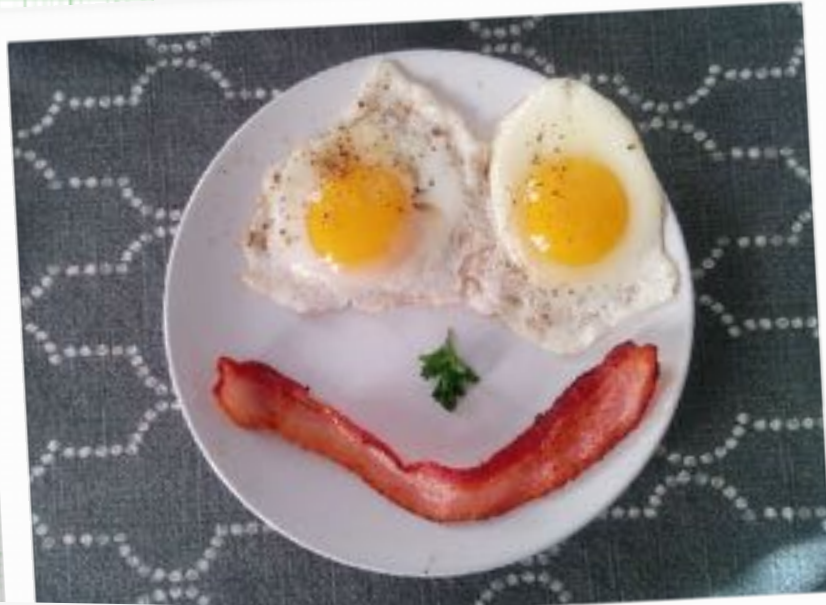
```
numCalories(breakfast)
```


VoxPL also lets you compose estimates

`numCalories(breakfast)`

+

`numCalories(lunch)`



VoxPL also lets you compose estimates

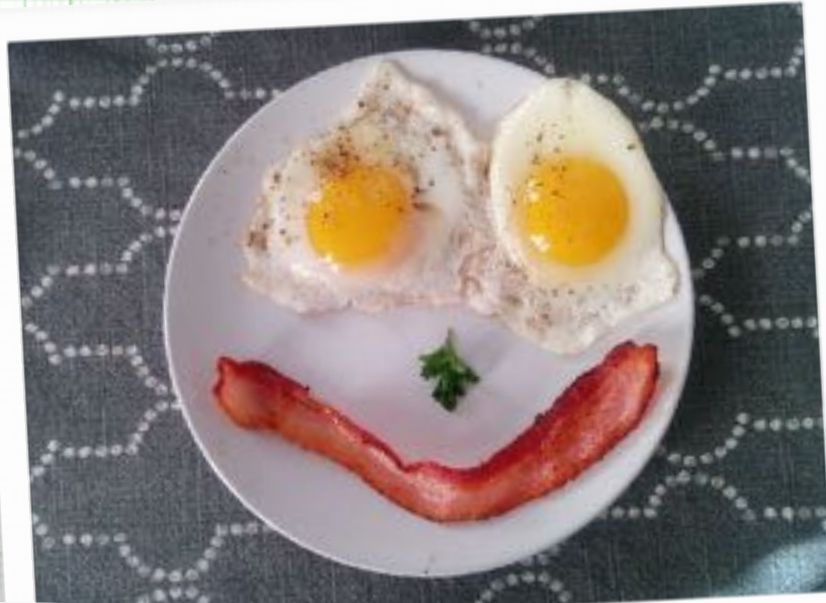
`numCalories(breakfast)`

+

`numCalories(lunch)`

+

`numCalories(dinner)`





VoxPL also lets you compose estimates

`numCalories(breakfast)`

+

`numCalories(lunch)`

+

`numCalories(dinner)`

(how? bootstrap)



Calorie Counter

How good are VoxPL's estimates?



- 208 images of school lunches w/ground truth kcal
- VoxPL: desired ± 50 kcal
actual mean error: ± 51.5 kcal
- Competitive with professional nutritionists!

Calorie Counter

How cheap are the estimates?



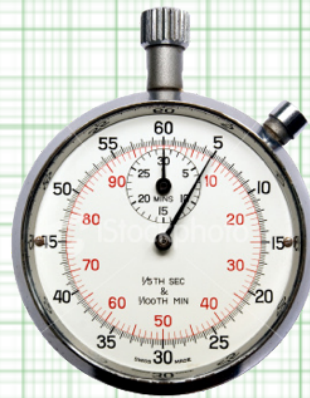
VoxPL:

± 50 cost: \$1.28

Nutritionist:

± 50 cost: \$100-200/hr

Programming with the Wisdom of the Crowd



Harnessing the crowd to do estimates (+ labeling)
with automatic budgeting, scheduling & quality control

<http://automan-lang.org>