Programming with People

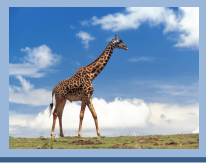
Emery Berger

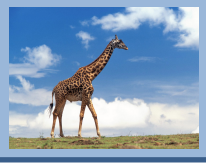
University of Massachusetts Amherst

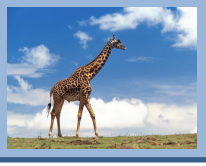

(science)

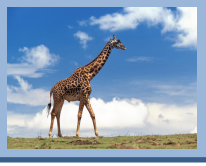

“is this a giraffe?”

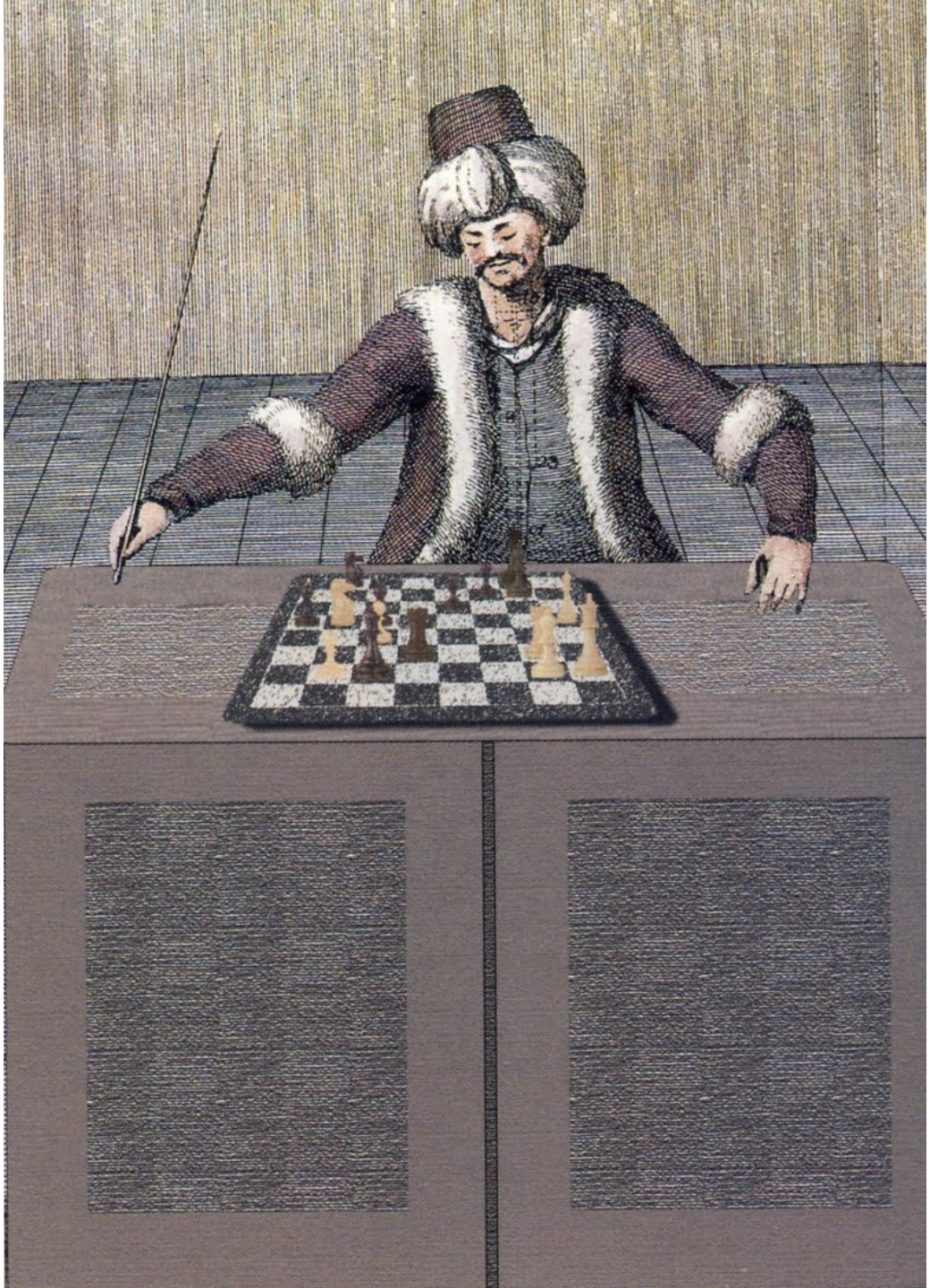

isGiraffe(*image*)

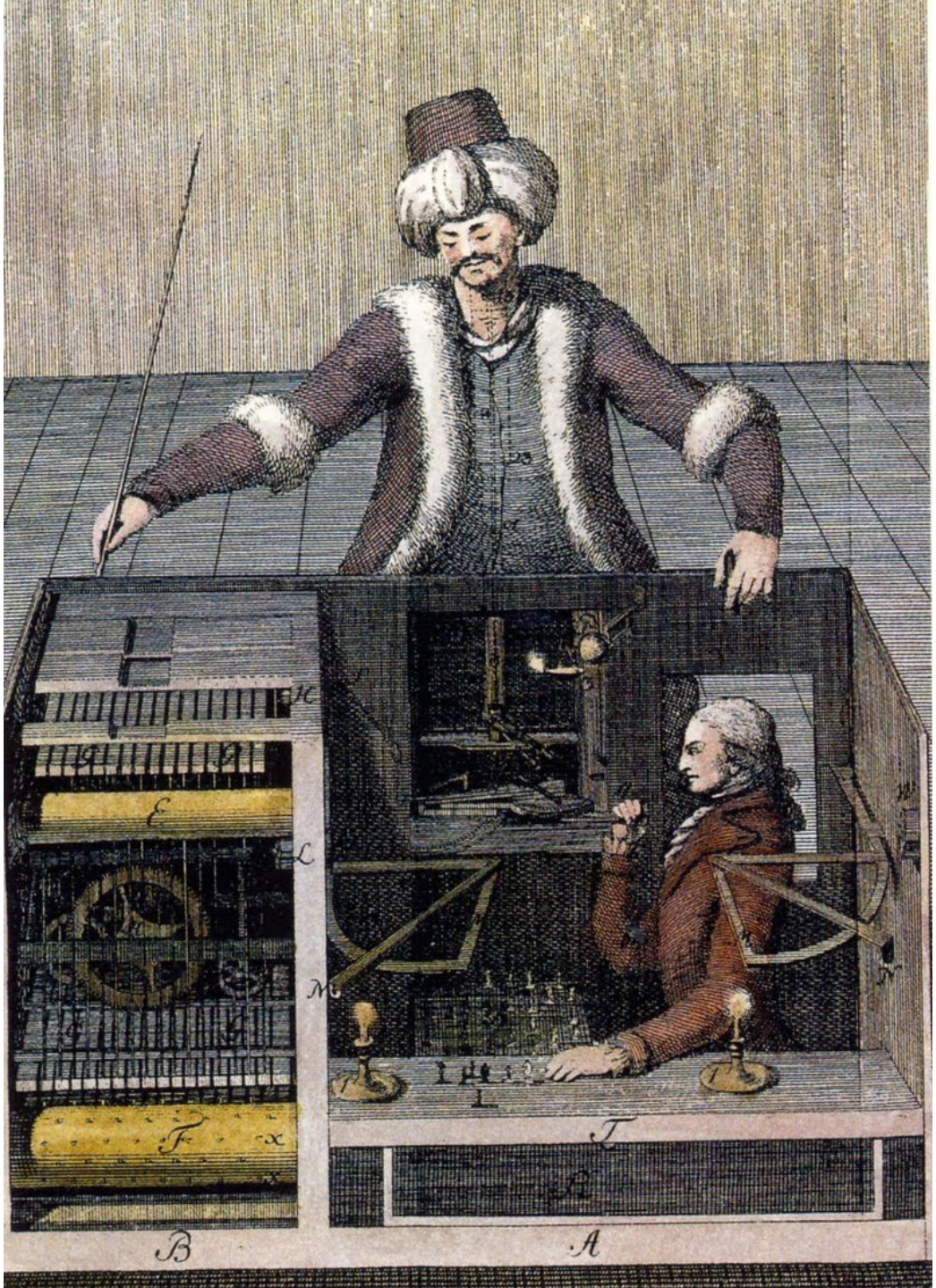
isGiraffe()

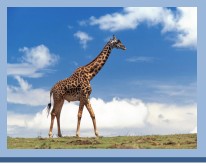
isGiraffe()

isGiraffe()

isGiraffe()




E
L
M
F
x
T
B
A

isGiraffe()

True

Timer: 00:00:00 of 10 minutes

Want to work on this HIT? Accept HIT

Want to see other HITs? Skip HIT

Total Earned: $0.12
Total HITs Submitted: 2

Find the Product Name (US-EN)

Requester: Classify AdImages

Reward: $0.06 per HIT

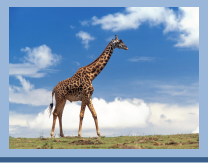
Qualifications Required: Tax Matters -8589092352025960714 is not less than 100, Product Name (US - EN) - Basic -858909241753507025

Please enter the name of the product being advertised

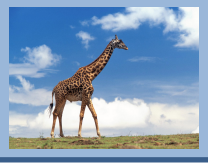
Complete These Steps

Please enter the name of the **product** being advertised

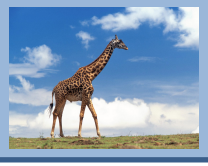
- This is not an ad
- The image failed to load
- The Product or Company Name is not known
- Contains adult content

isGiraffe()

True

isGiraffe()

True

isGiraffe()

False

False

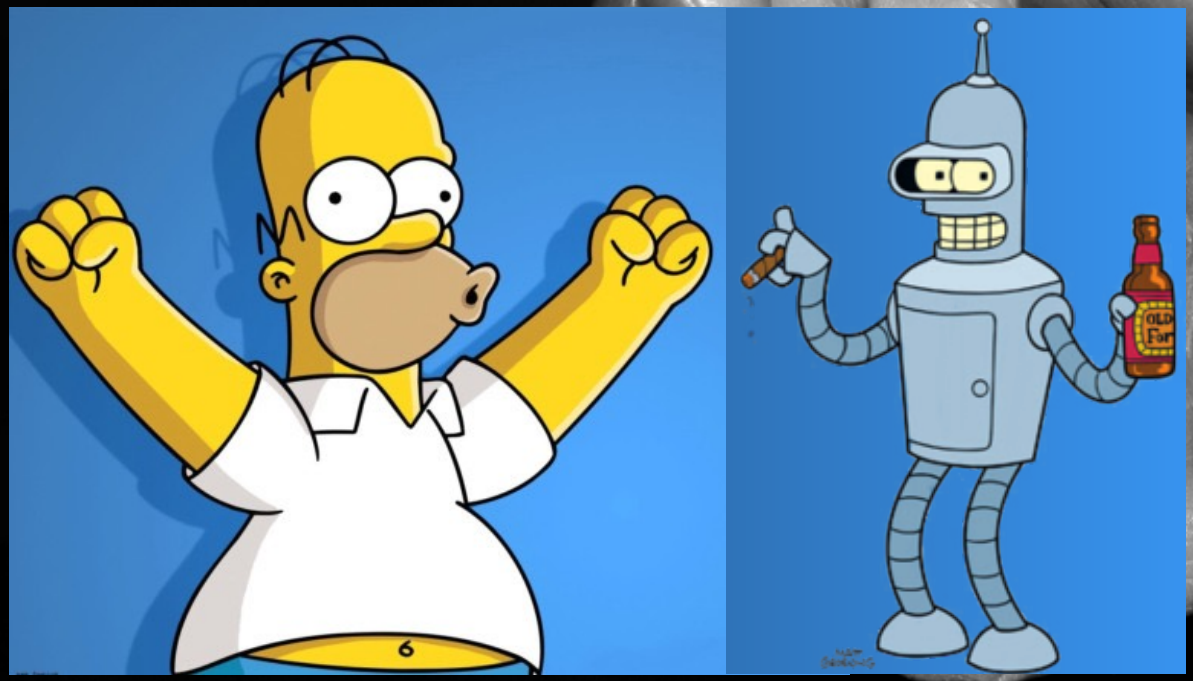
False

False

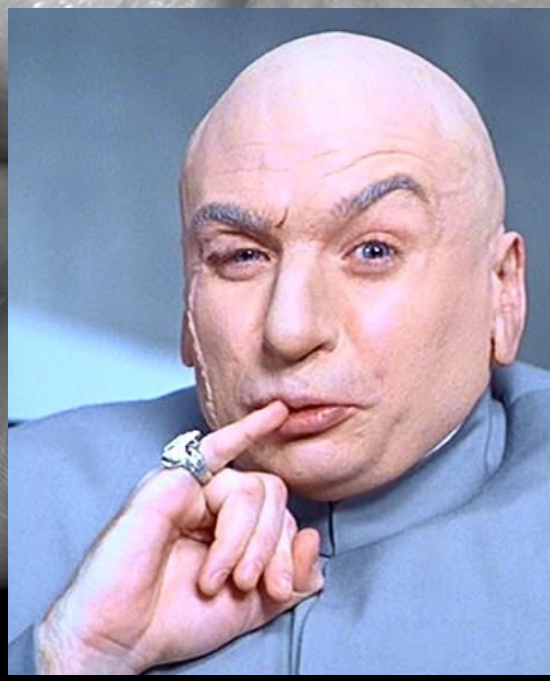
False


random adversary model

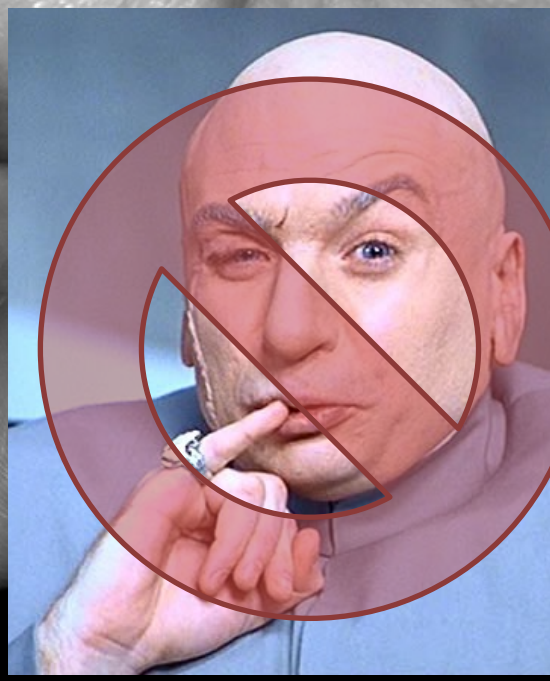
random adversary model



random adversary model



random adversary model



Qualifications Required:

HIT abandonment rate (%) is less than 20

HIT approval rate (%) is greater than 95

Qualifications Required:

Total approved HITs is greater than 1000

Categorizati

HIT approva

Qualifications Required:

Total approved HITs is greater than 1000

HIT approval rate (%) is not less than 95

Qualifications Required:

Total approved HITs is greater than 2500

HIT approval rate (%) is greater than 95

Location is US

long-term financial incentive

Qualifications Required:
HIT abandonment rate (%) is less than 20
HIT approval rate (%) is greater than 95

Qualifications Required:
Total approved HITs is greater than 1000
Categorizati
HIT approva

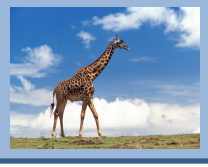
Qualifications Required:
Total approved HITs is greater than 1000
HIT approval rate (%) is not less than 95

Qualifications Required:
Total approved HITs is greater than 2500
Location is US

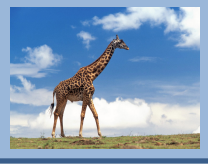
credentials limit *Sybil* attacks

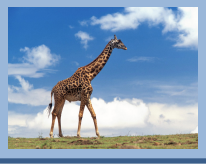
Step 3: Get paid for your work

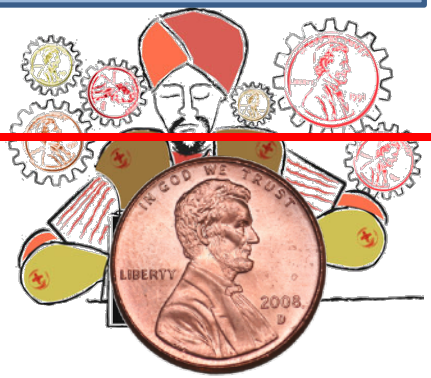
After the requester approves your work, money is deposited into your Amazon Payments account.

isGiraffe()

?

isGiraffe()

isGiraffe()



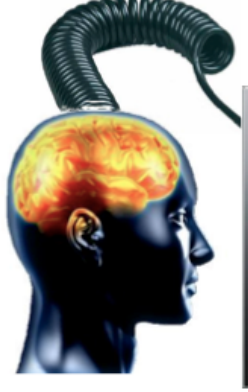
Pr[agree] = 1/2

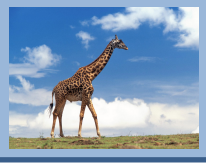
Pr[agree] = 1/32

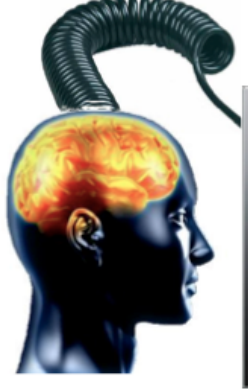
< 5%

$$\Pr[\text{agree}] = k(1/k)^n$$

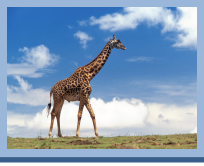
(unanimous agreement)

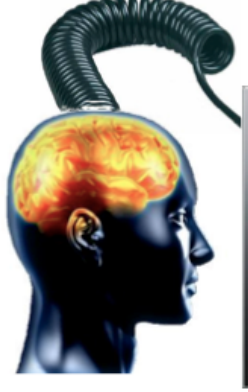


isGiraffe()



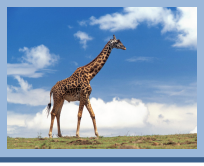
(Scala)

isGiraffe()



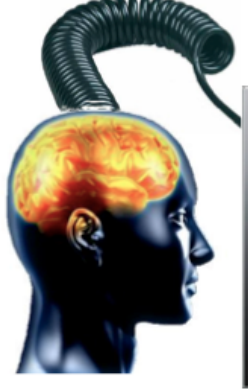
AUTOMAN

(Scala)

isGiraffe()

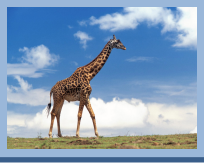
AutoMan programmer-specified

Total $ for computation



AUTOMAN

(Scala)

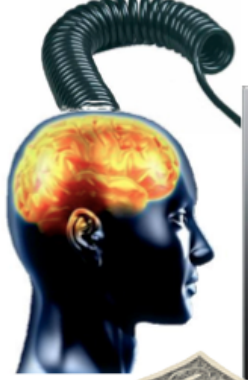
isGiraffe()

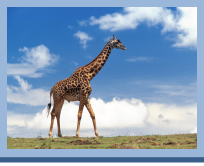
AutoMan programmer-specified

Total $ for computation

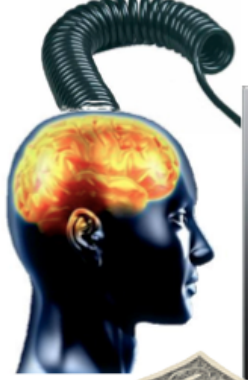
95%
(p < 0.05)

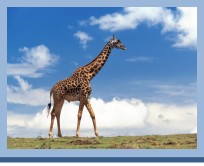
Confidence level
(per function)



isGiraffe()

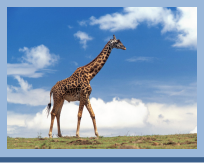
US minimum wage,
Adaptive doubling



isGiraffe()

US minimum wage,
Adaptive doubling

30s, $0.06 ($7.25 / 120)

isGiraffe()

US minimum wage,
Adaptive doubling
prevents gaming

~~30s, $0.06 ($7.25 / 120)~~

60s, $0.12


amazon.com


5 EURO


amazon.com


10
10 EURO


amazon.com


20
20 EURO


amazon.com


20
20 EURO


amazon.com


20
20 EURO


amazon.com


amazon.com

$$E[gain] = base\ (P_{avail})^{round} * multiplier^{round}$$

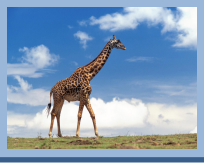
E[gain]= base $(½)^{round}$ * $multiplier^{round}$

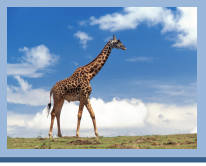
E[gain]= base $(½)^{round}$ * 2^{round}

E[gain]= base

E[gain]= base

no incentive to wait

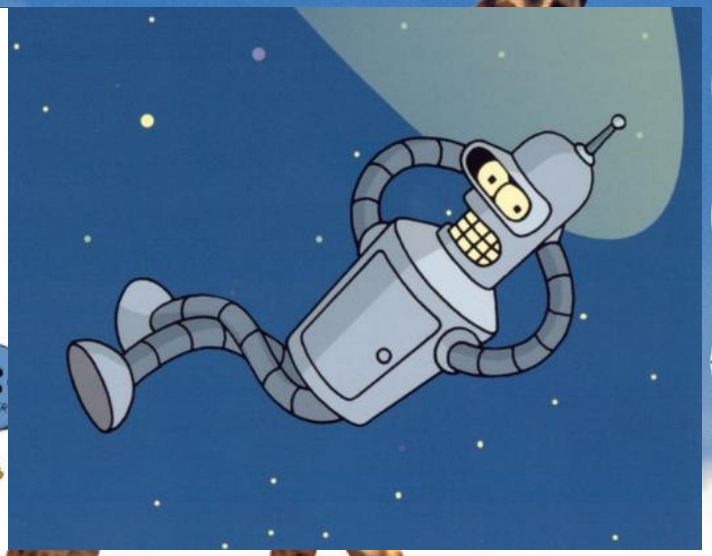
isGiraffe()

isGiraffe()

True *95% confidence*

How many giraffes are in this picture?

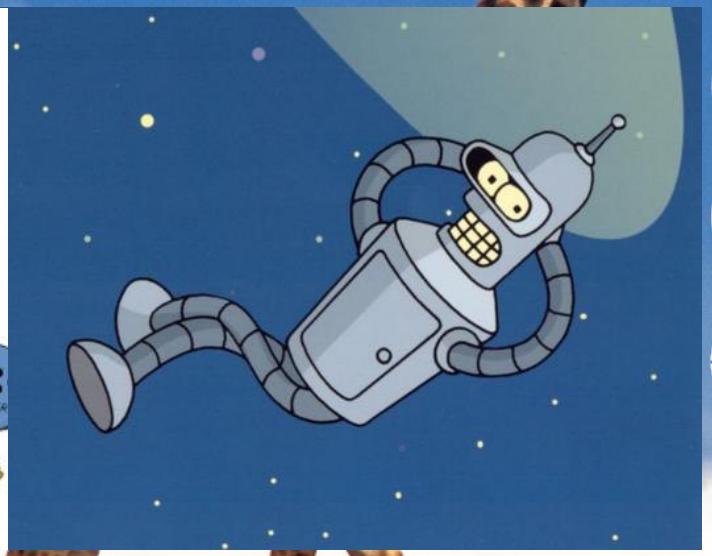
How many giraffes are in this picture?



- None
- More than one
- One

k choices

How many giraffes are in this picture?



- None
- More than one
- One

k choices

How many giraffes are in this picture?

How many giraffes are in this picture?

Which are from Sesame Street?

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

2^k choices

Which are from Sesame Street?

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

2^k choices

Which are from Sesame Street?

- [x] Oscar the Grouch
- [ ] Kermit the Frog
- [x] Spongebob Squarepants
- [x] Cookie Monster
- [ ] The Count

2^k choices

What does this license plate say?

What characters are printed on this license plate?

XXXXXX

[A-Z0-9]{6}

36^6 choices = 2176782336

Which one of these doesn't belong?

[95% conf.]

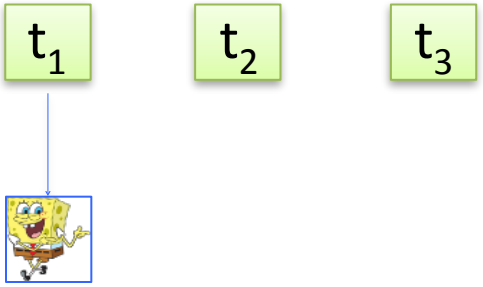
AUTOMAN: spawns 3 tasks @ $0.06; 30s work

t_1 t_2 t_3

Which one of these doesn't belong?

[95% conf.]

AUTOMAN: spawns 3 tasks @ $0.06; 30s work

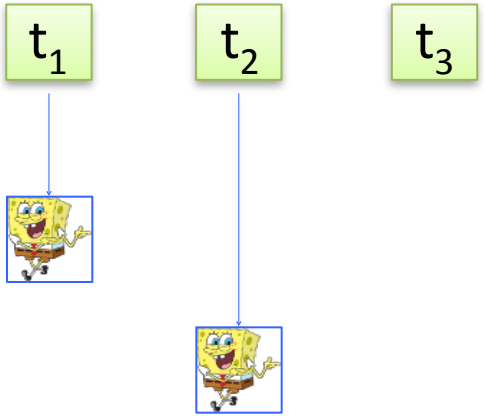


1m 50s

Which one of these doesn't belong?

[95% conf.]

AUTOMAN: spawns 3 tasks @ $0.06; 30s work



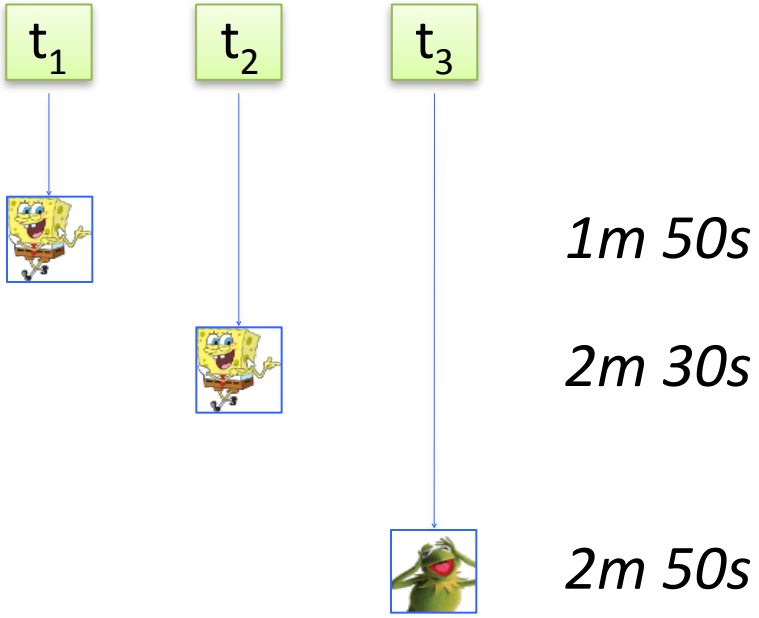
1m 50s

2m 30s

Which one of these doesn't belong?

[95% conf.]

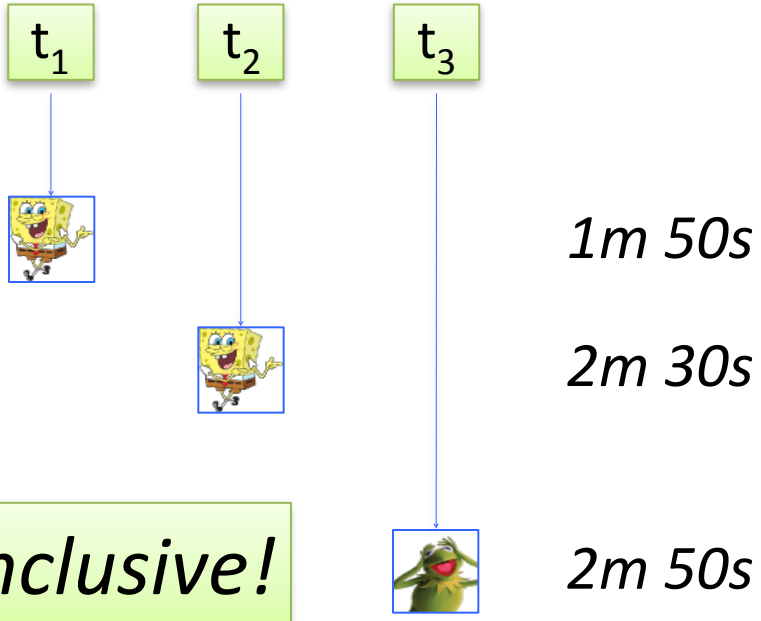
AUTOMAN: spawns 3 tasks @ $0.06; 30s work



Which one of these doesn't belong?

[95% conf.]

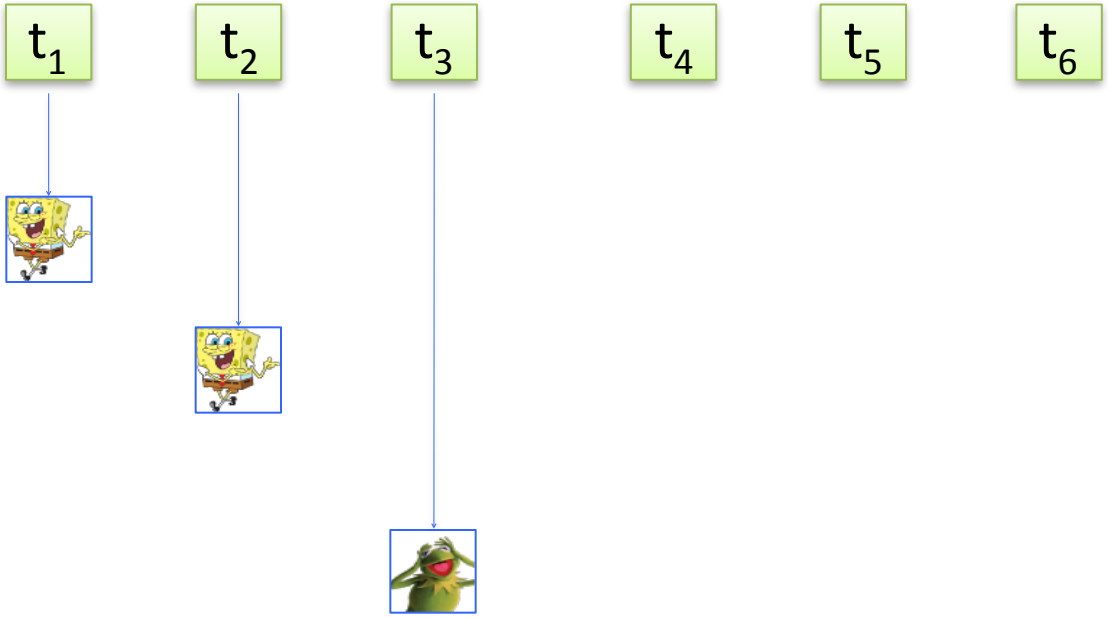
AUTOMAN: spawns 3 tasks @ $0.06; 30s work



Which one of these doesn't belong?

[95% conf.]

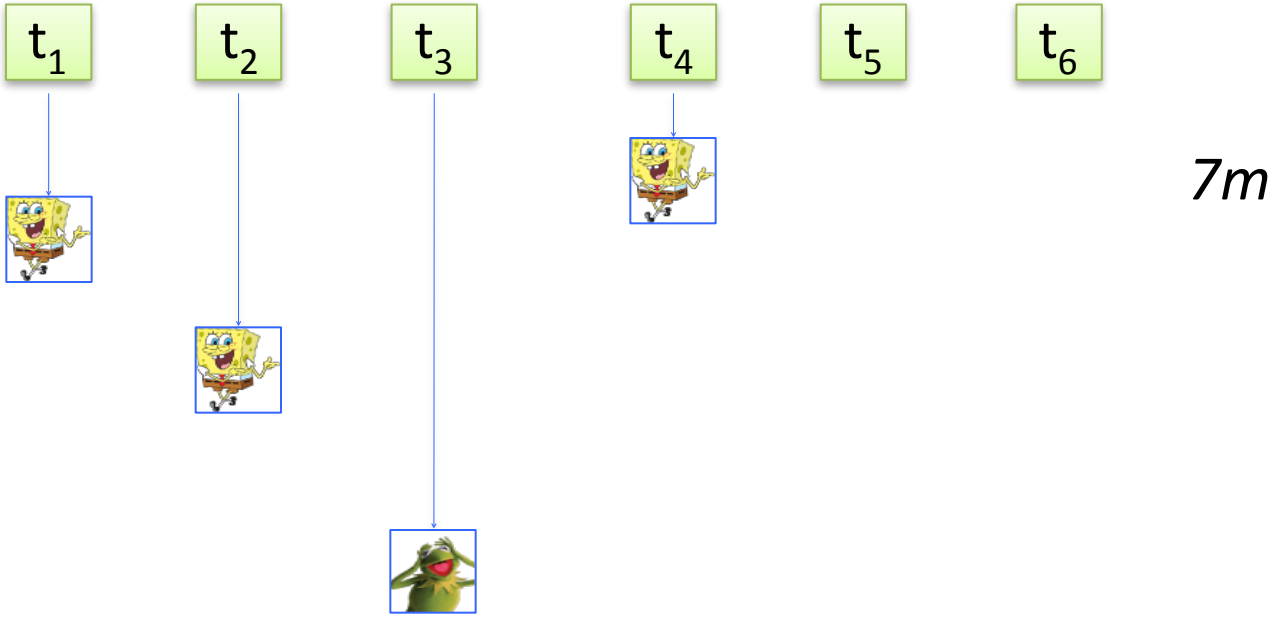
AUTOMAN: spawns 3 more tasks



Which one of these doesn't belong?

[95% conf.]

AUTOMAN: spawns 3 more tasks

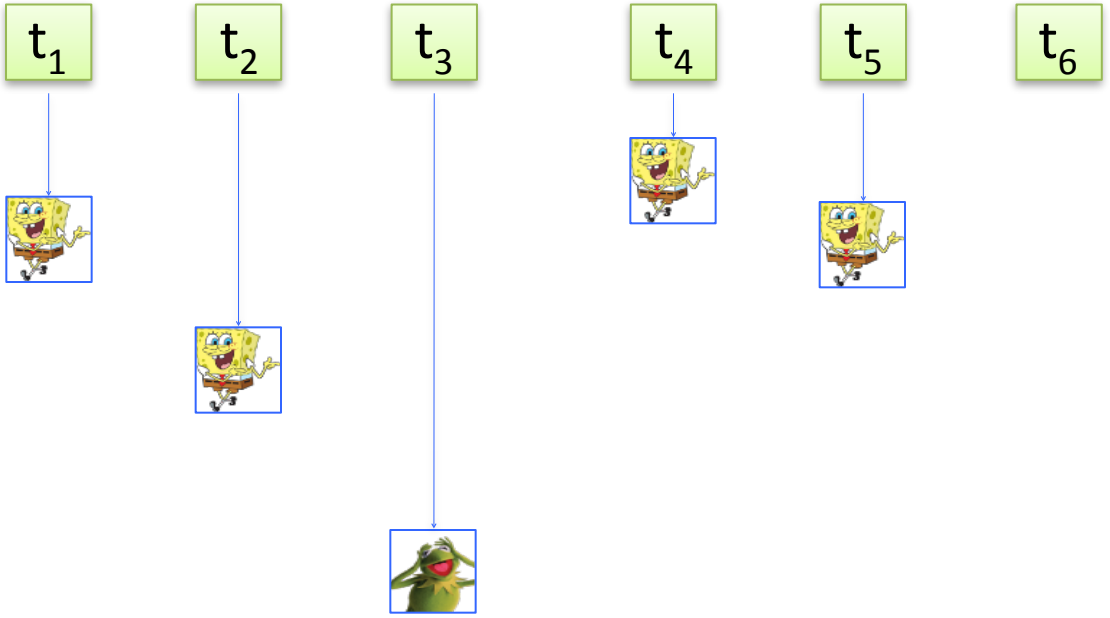


7m

Which one of these doesn't belong?

[95% conf.]

AUTOMAN: spawns 3 more tasks



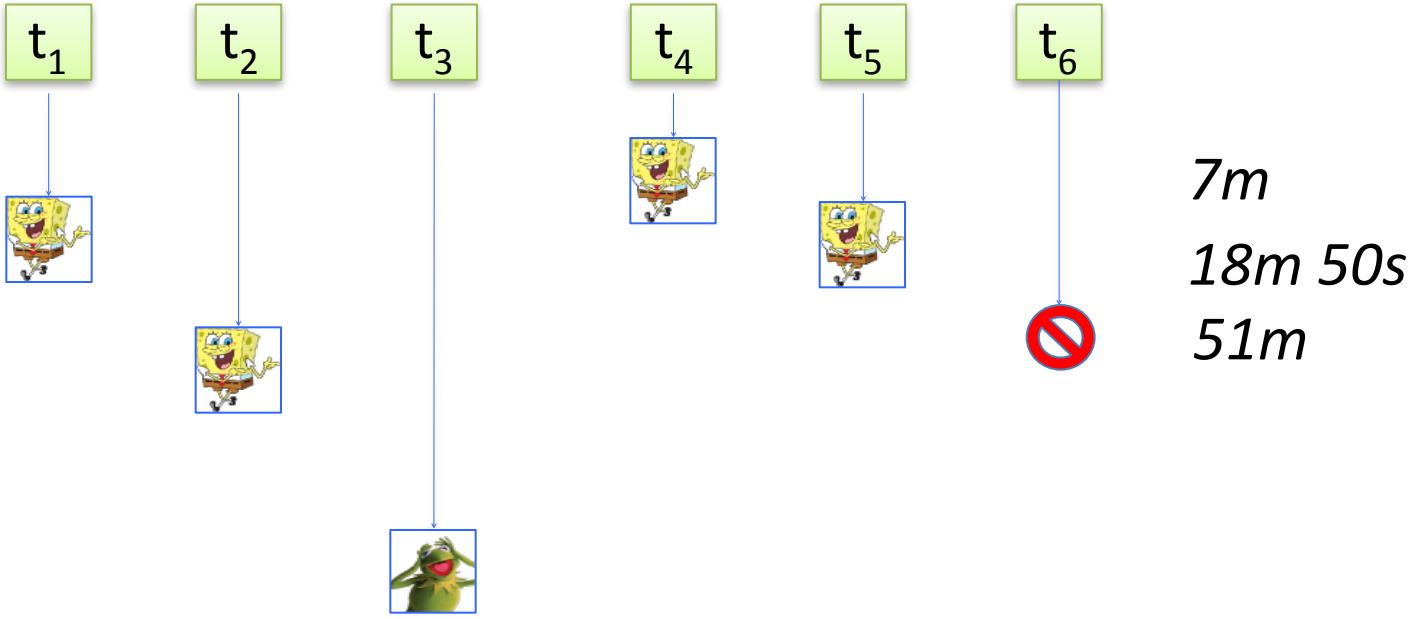
7m

18m 50s

Which one of these doesn't belong?

[95% conf.]

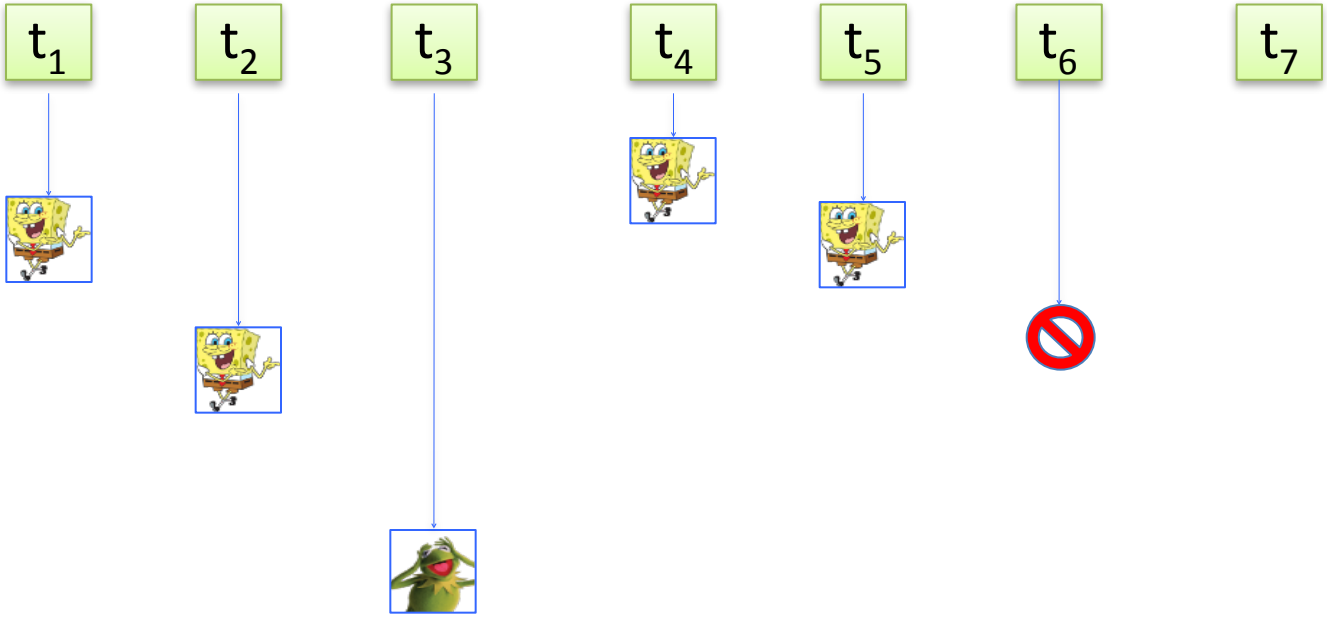
AUTOMAN: spawns 3 more tasks



Which one of these doesn't belong?

[95% conf.]

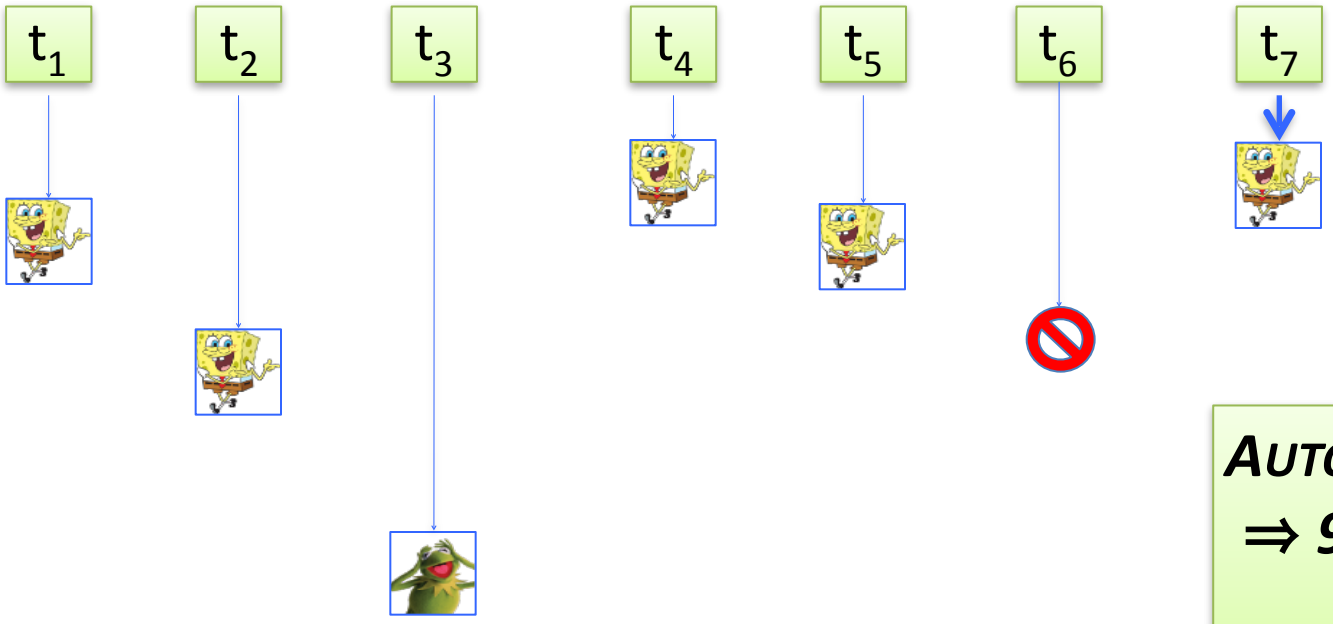
AUTOMAN: spawn 1 more task @ $0.12, 60s work



Which one of these doesn't belong?

[95% conf.]

AUTOMAN: spawn 1 more task @ $0.12, 60s work



1h 9m 50s;
cost = $0.36

AUTOMAN: 5 out of 6
⇒ 95% confidence;
return 


read_plate()
767 JKF


GIFTS & HERBS
ONE WAY
JACKSON
TAXI
VOLVO
5LHJ609


Massachusetts
10
BOSTON RED SOX
JF 1
The Jimmy Fund


Massachusetts
RW0000
Preserve the Trust


California
USA
SAMPLE


California
SAM123


California
PROTECT OUR COAST & OCEAN


California
SAM123


California
3SAM123
PROTECT LAKE TAHOE


CALIFORNIA
SAM123
WE WILL NEVER FORGET


California
3SAM123
YOSEMITE NATIONAL PARK


California
UCLA
SAM123


APR
Massachusetts
CI0000
Cape Cod & Islands


Massachusetts
14
UMASS
UM 1345
You were. You are. UMASS.


165M575


MXH 979


40L-190


ZPP-626

Timer: 00:00:05 of 30 seconds

Finished with this HIT? Submit HIT

Let someone else do it? Return HIT

Automatically accept the next HIT

Total Earned: $0.77
Total HITs Submitted: 17

Recognize this license plate.

Requester: Dan Barowy **Reward:** $0.06 per HIT **HITs Available:** 5 **Duration:** 30 seconds

Qualifications Required: None

What are the characters printed on this license plate? (CAPS ONLY, NO DASHES, DOTS OR SPACES please!)

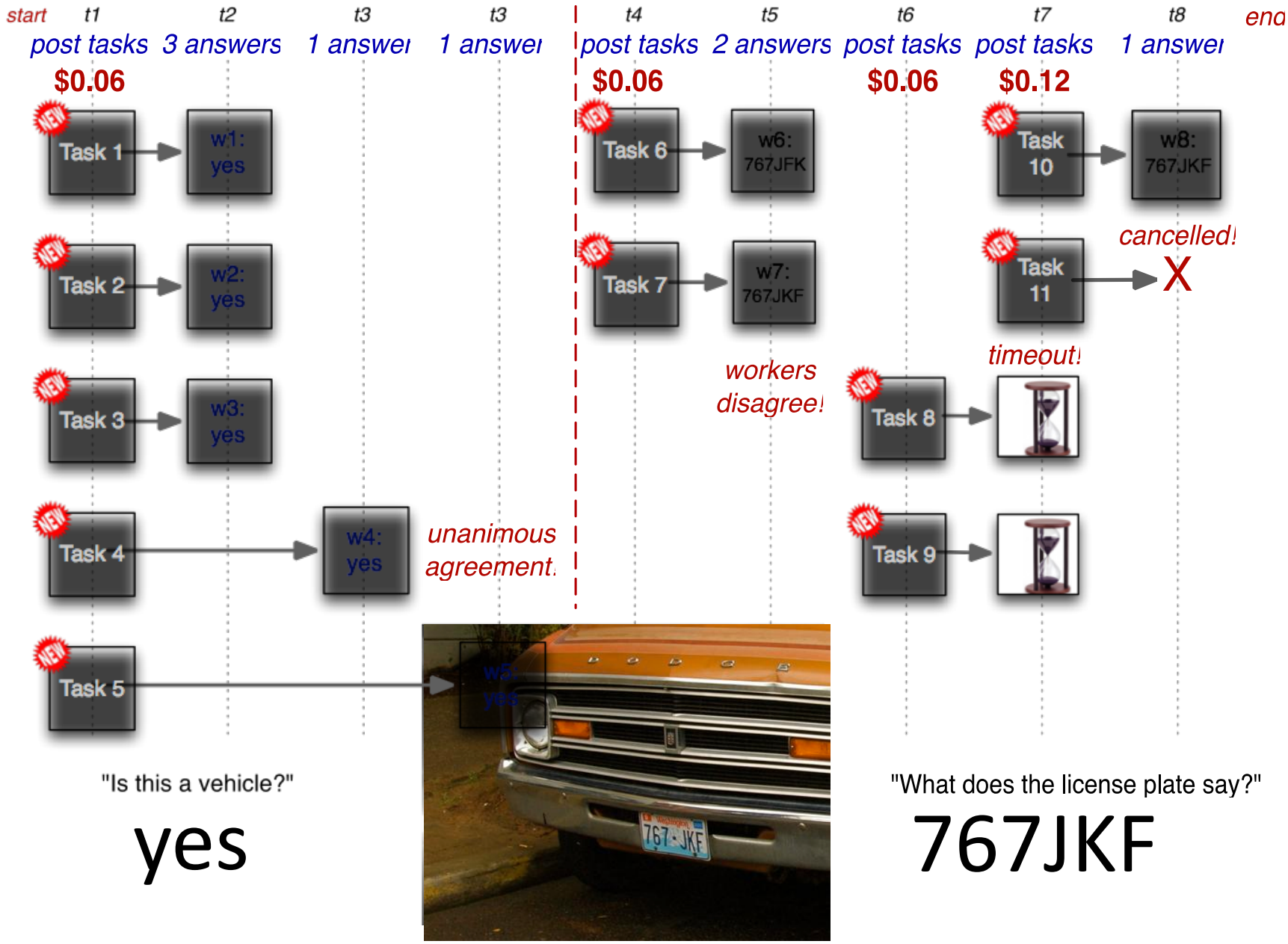
```
def is_car(img_url: String) =
  a.RadioButtonQuestion { q =>
    q.budget = 1.00
    q.confidence = 0.95
    q.text = “Is this a car?”
    q.image_url = img_url
    q.options = List(
      a.Option('yes, ”Yes"),
      a.Option('no, ”No”)
    )
  }
```

```
def get_plate_text(img_url: String) =
   a.FreeTextQuestion { q =>
      q.text = ”What does this plate
                  say?"
      q.image_url = img_url
      q.pattern = "XXXXXYYY”
}
```

```
def get_plate_text(img_url: String) =
  a.FreeTextQuestion { q =>
    q.text = ”What does this plate
                say?"
    q.image_url = img_url
    q.pattern = "XXXXXYYY”
}
```

```
val plate_texts = s3_urls.par.map
{ url => get_plate_text(url) }

plate_texts.foreach { text =>
println(text) }
```


start
t1
t2
t3
t3
t4
t5
t6
t7
t8
end
post tasks
3 answers
1 answer
1 answer
post tasks
2 answers
post tasks
post tasks
1 answer
$0.06
$0.06
$0.06
$0.12
Task 1
w1: yes
Task 2
w2: yes
Task 3
w3: yes
Task 4
w4: yes
Task 5
w5: yes
Task 6
w6: 767JFK
Task 7
w7: 767JKF
Task 10
w8: 767JKF
Task 11
cancelled!
X
timeout!
workers disagree!
Task 8
Task 9
unanimous agreement!
"Is this a vehicle?"
yes
"What does the license plate say?"
767JKF


VOLVO
F10
PNN-3715


F12
INTERCOOLER
PNP-4416


F12
INTERCOOLER
PNP-4416


PNP-4415


57642


VOLVO


F12
PNP-4141

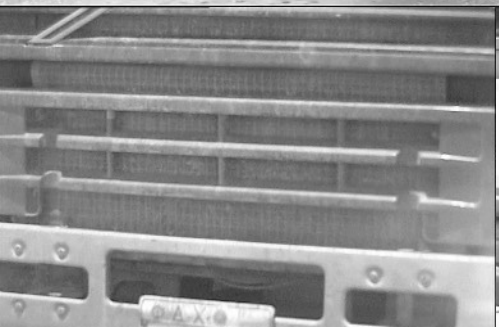

VOLVO


ABS
ASR
EPS


INTERCOOLER
SCANIA


143M
450
SCANIA
PNP-4411


XHO-8584




PNN-3885


PNN-3715


PNP-4236
VOLVO

MediaLab LPR database

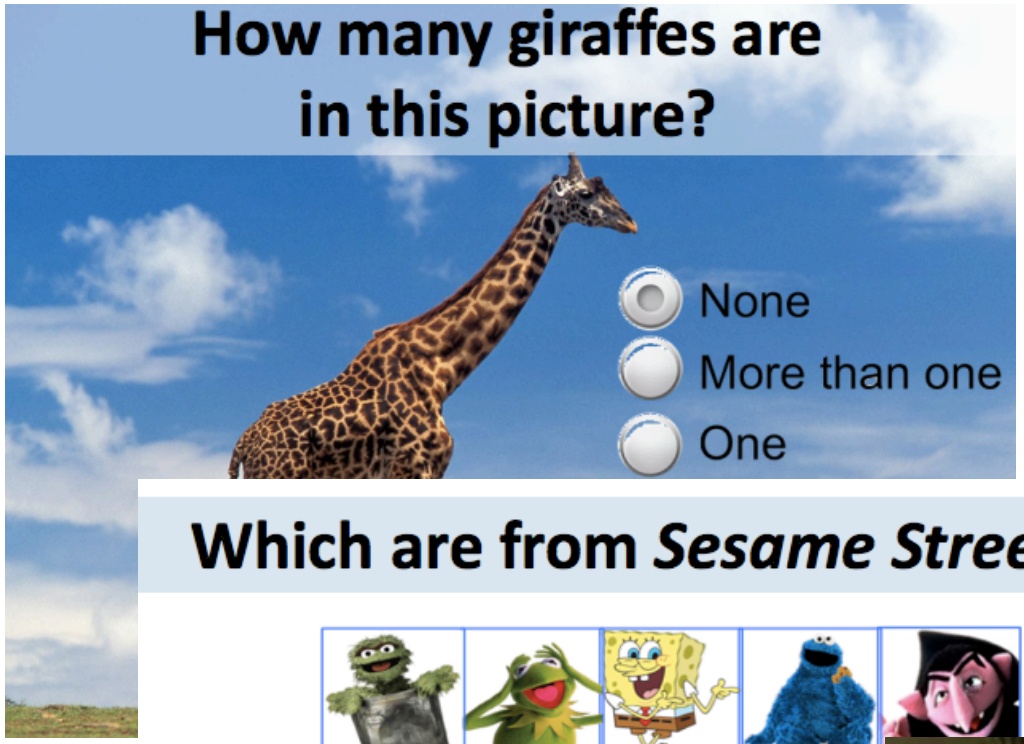
"extremely difficult" dataset

144 plates

Accuracy: 91.6% >12.2%!

Average cost: 12.08 cents

Latency: < 2 minutes per image

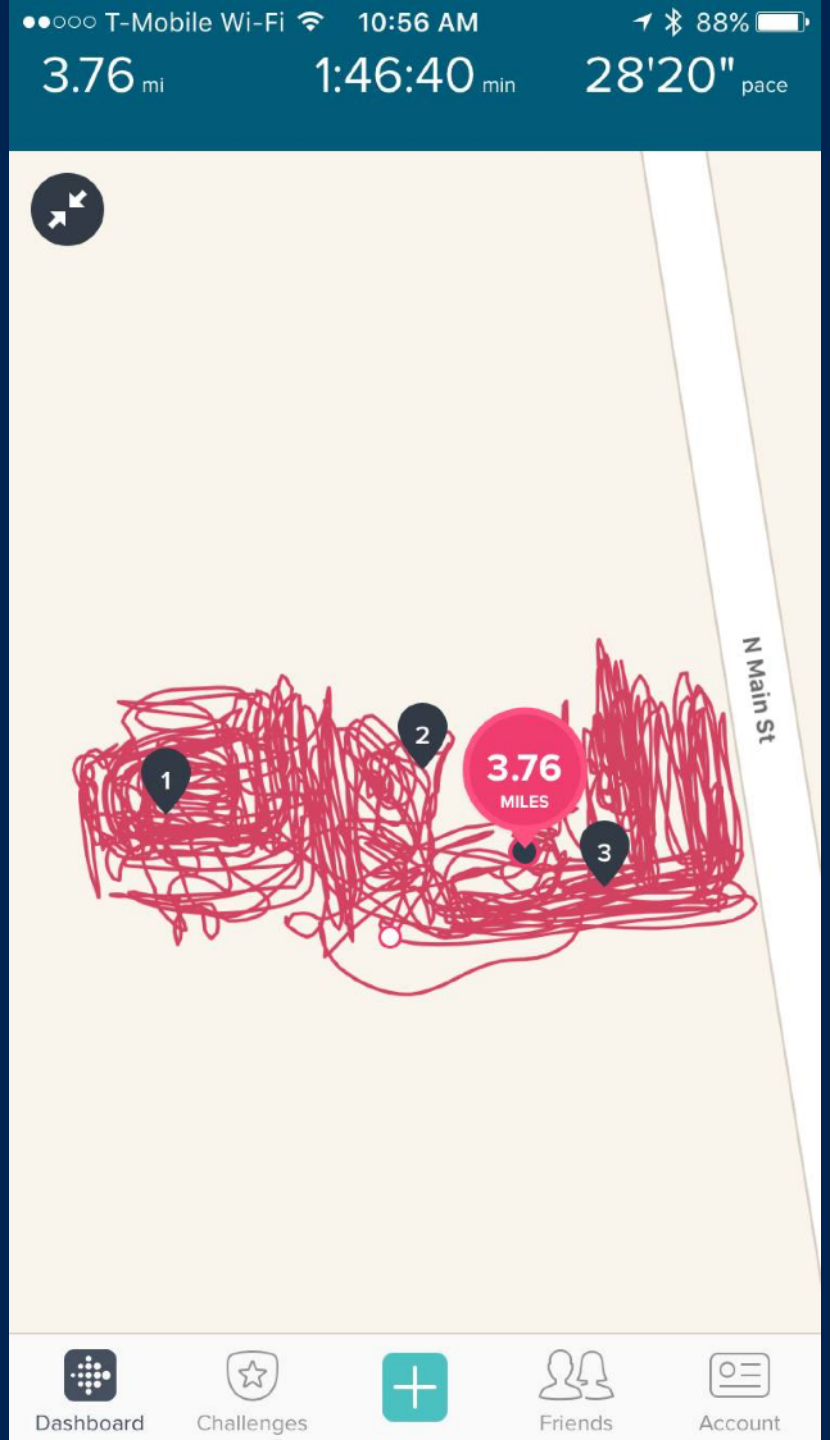


Etsy

Which are from *Sesame Street*?

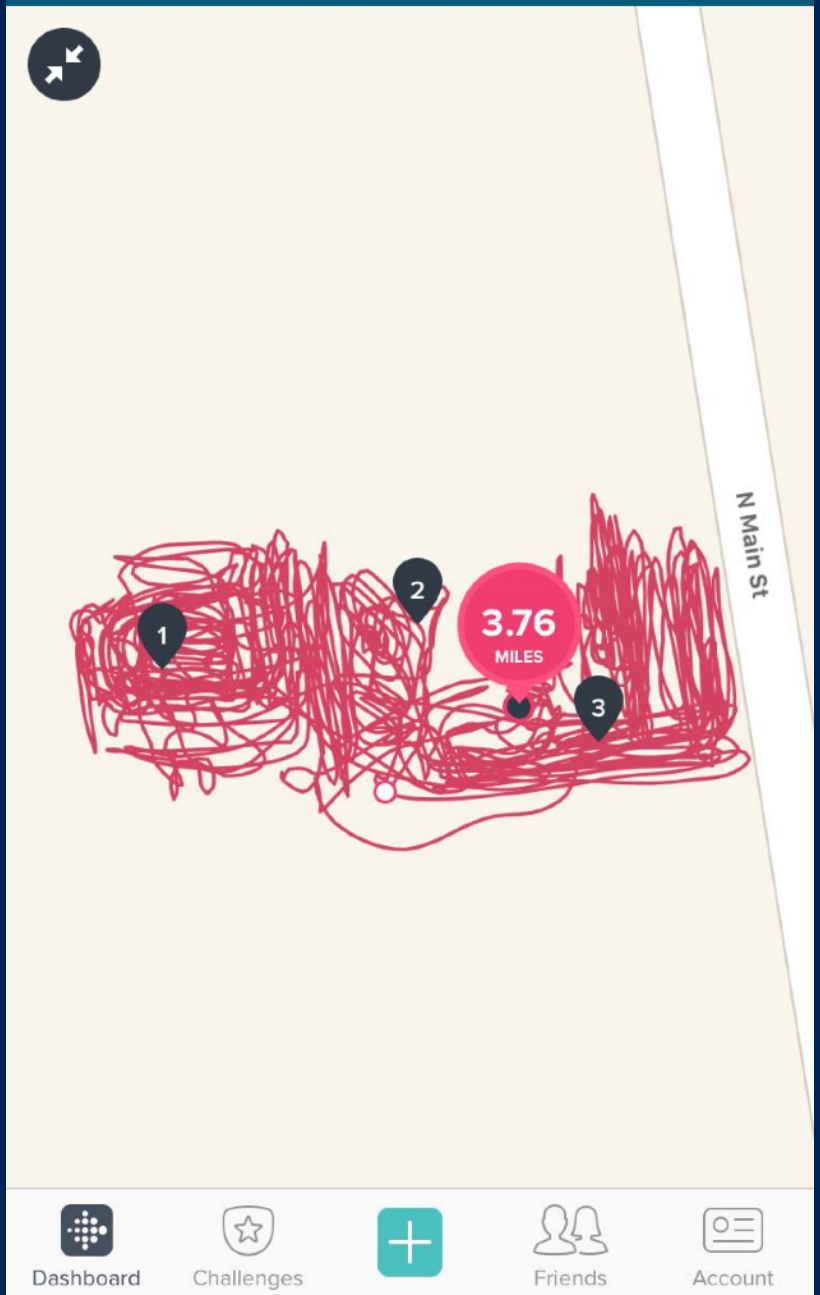
- [x] Oscar the Grouch
- [ ] Kermit the Frog
- [x] Spongebob Squarepants
- [x] Cookie Monster
- [ ] The Count


12:58
86
fitbit


T-Mobile Wi-Fi
10:56 AM
88%
3.76 mi
1:46:40 min
28'20" pace
N Main St
1
2
3.76
MILES
3
Dashboard
Challenges
Friends
Account

●●○○○ T-Mobile Wi-Fi 10:56 AM 88%

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



Dashboard Challenges Friends Account

●●○○○ T-Mobile Wi-Fi 4:03 PM 89%

Exercise

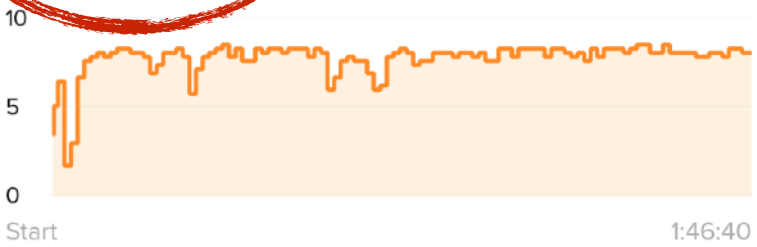
Hike

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

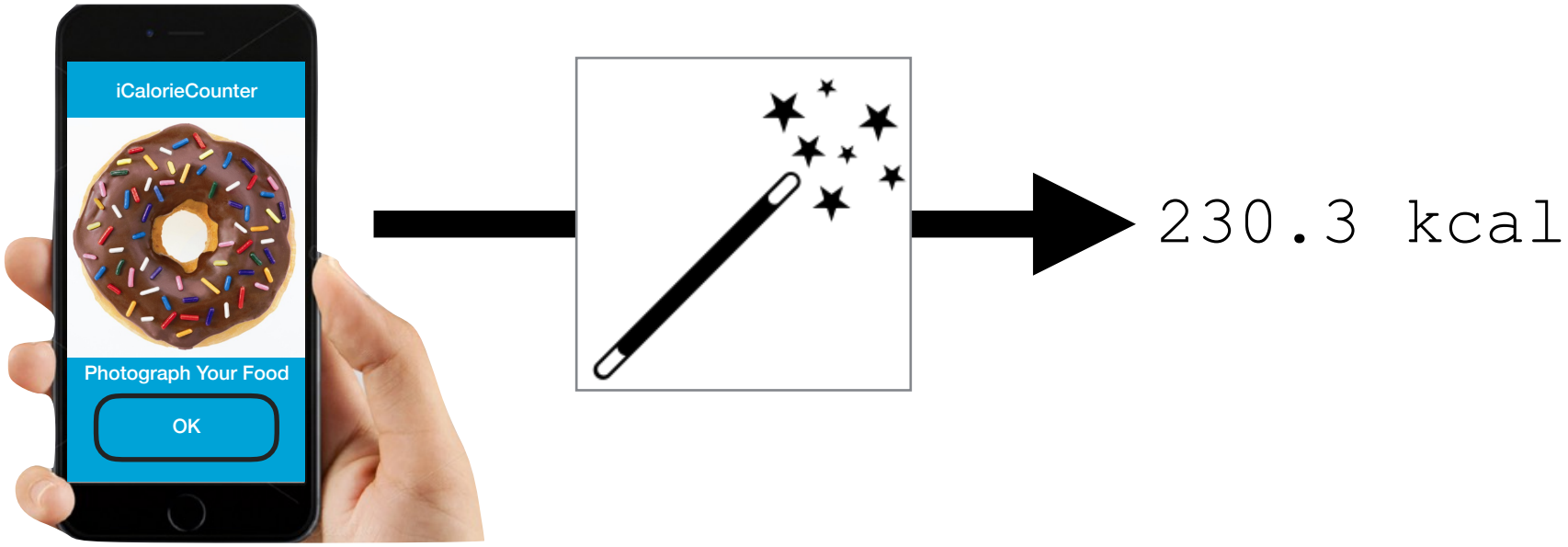
+11,599

of **14,529** steps taken

+835

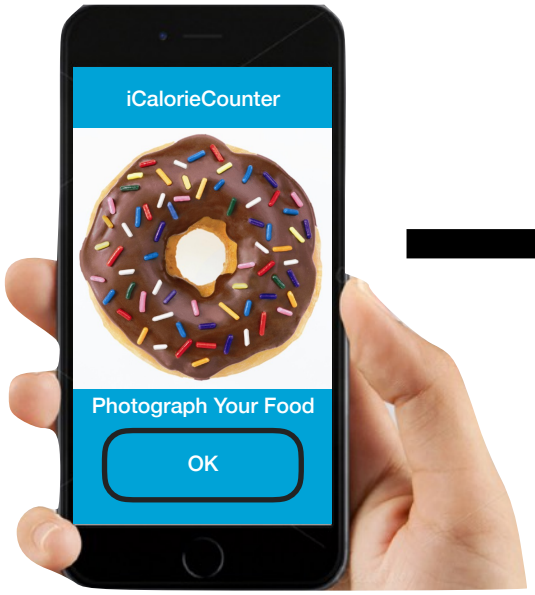
of **2,717** calories burned

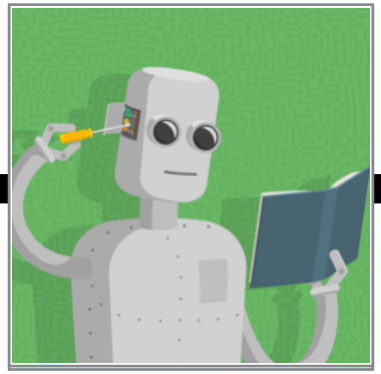
Dashboard Challenges Friends Account



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


INTRODUCTION TO
ALGORITHMS





230.3 kcal

1. take a photo

2. machine learning algorithms (AI)

3. return estimate

Im2Calories: towards an automated mobile vision food diary

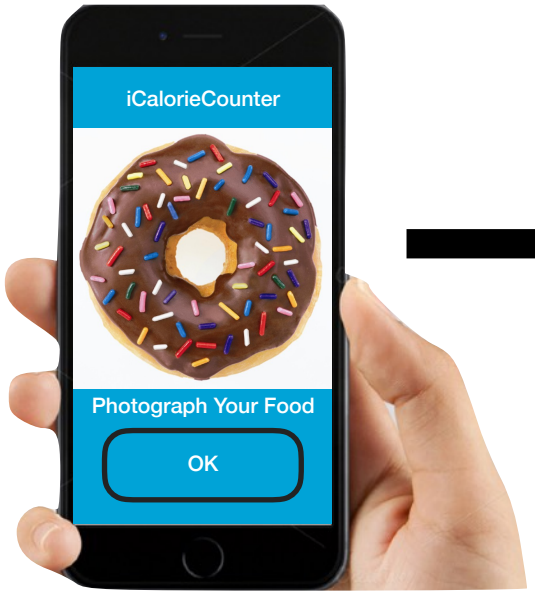
Once we have determined that the image contains a meal, we try to analyze its contents. The first step is to determine which restaurant the user is in. In this paper, we use Google's Places API [27] for this. We then retrieve the menu of the nearest restaurant from the web,[5] parse the menu of K food items, and retrieve images for

orban

evin Murphy

n

a system. Our approach utilizes several deep learn

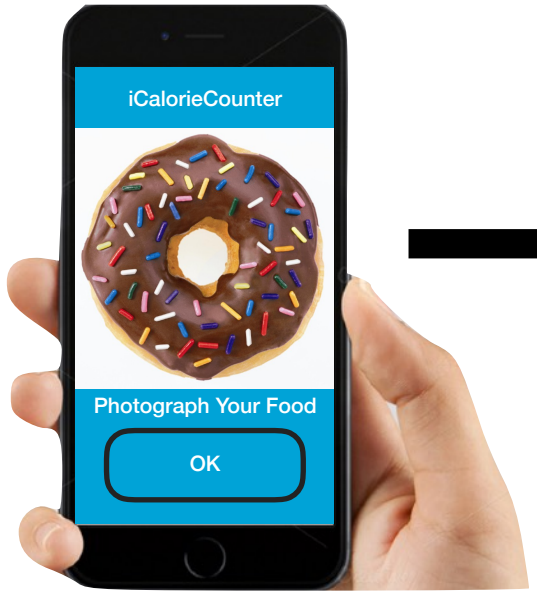


230.3 kcal

1. take a photo

2. machine learning

3. return estimate



Auto
Man

230.3 kcal

1. take a photo

2. crowdsourcing

2. human

3. return estimate

readPlate()

AutoMan

767-JKF ~~76-JFK~~ ~~AZHG6A~~ 767-JKF 767-JKF

767-JKF

"How many calories are in this donut?"

2

326

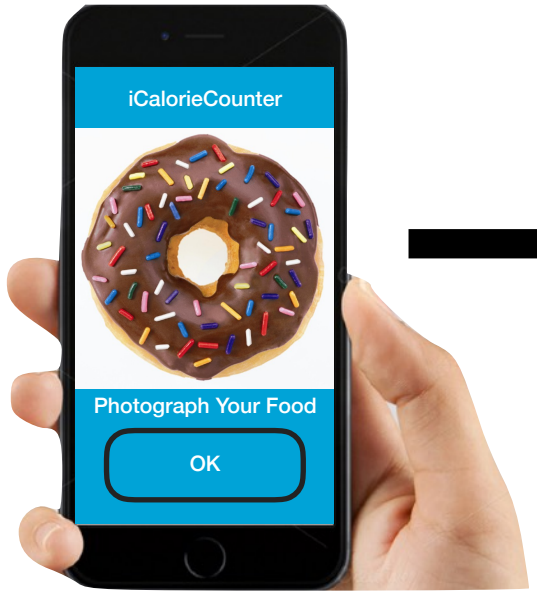
214

318

283

274

214 ≠ 318 ≠ 283 ≠ 274



VoxPL

230.3 kcal

1. take a photo

2. crowdsourcing

3. return estimate

VoxPL

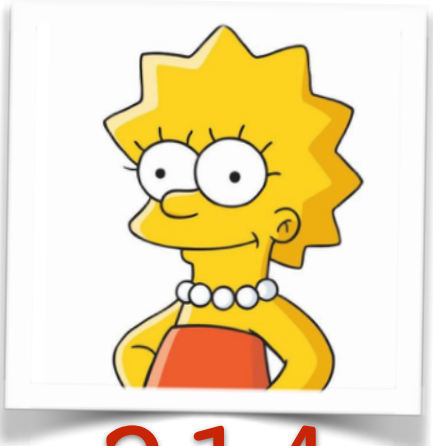
Extends AutoMan DSL with estimates

```
def numCalories(url: String) = Estimate (
  confidenceInterval = SymmetricCI(50),
  text = "How many calories are
          in the food pictured?",
  imageUrl = url
)
```

"How many calories are in this donut?"

2

326



214

318

283

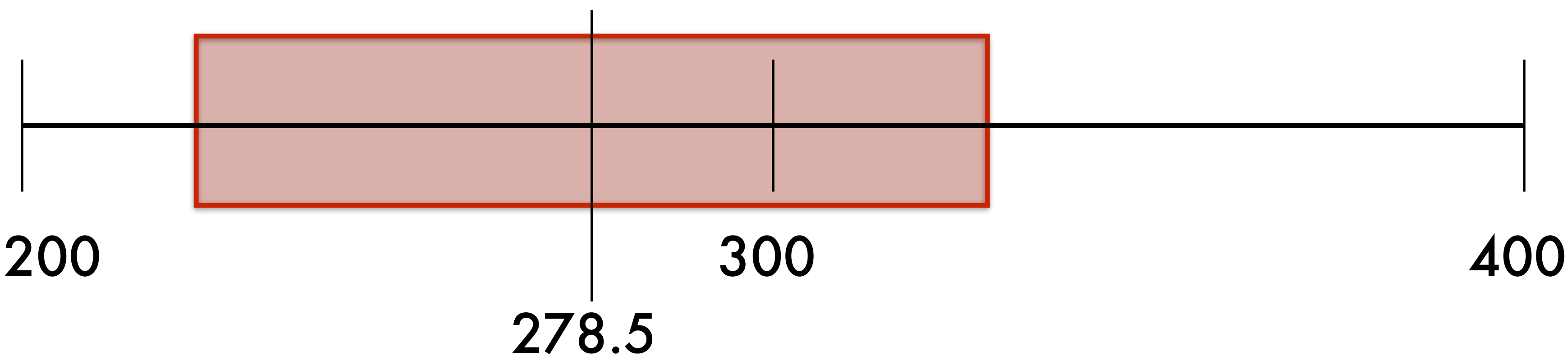
274

214 ≠ 288.5 ≠ 274

mean ≠ median ≠ mode

What would it take

to trust that median value of 278.5 is a good estimate?



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

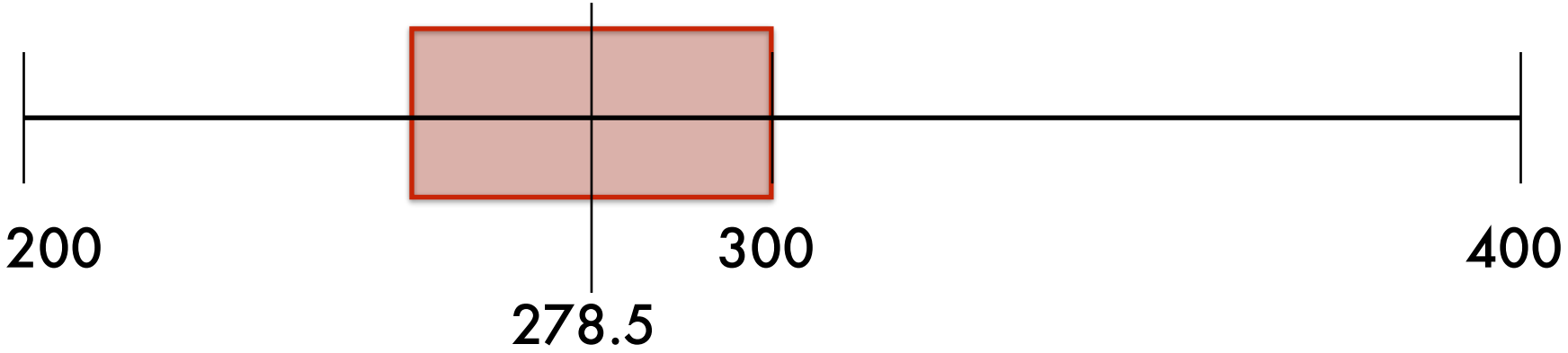
What about confidence intervals?

In other words, which opinions probably don't belong to Homers?

“Donut contains 278.5±50 kcal.”

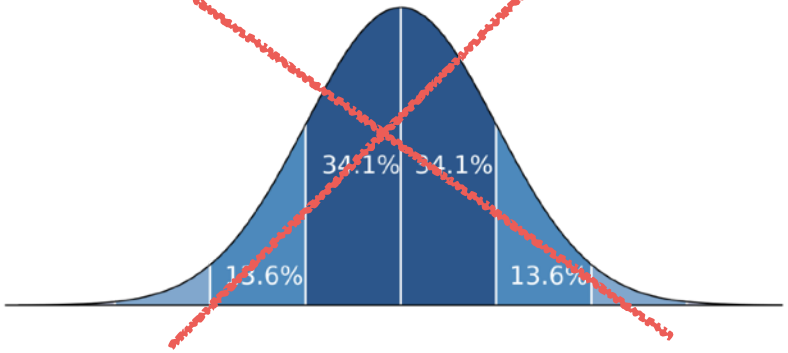
What would it take

to trust that median value of 278.5 is a good estimate?

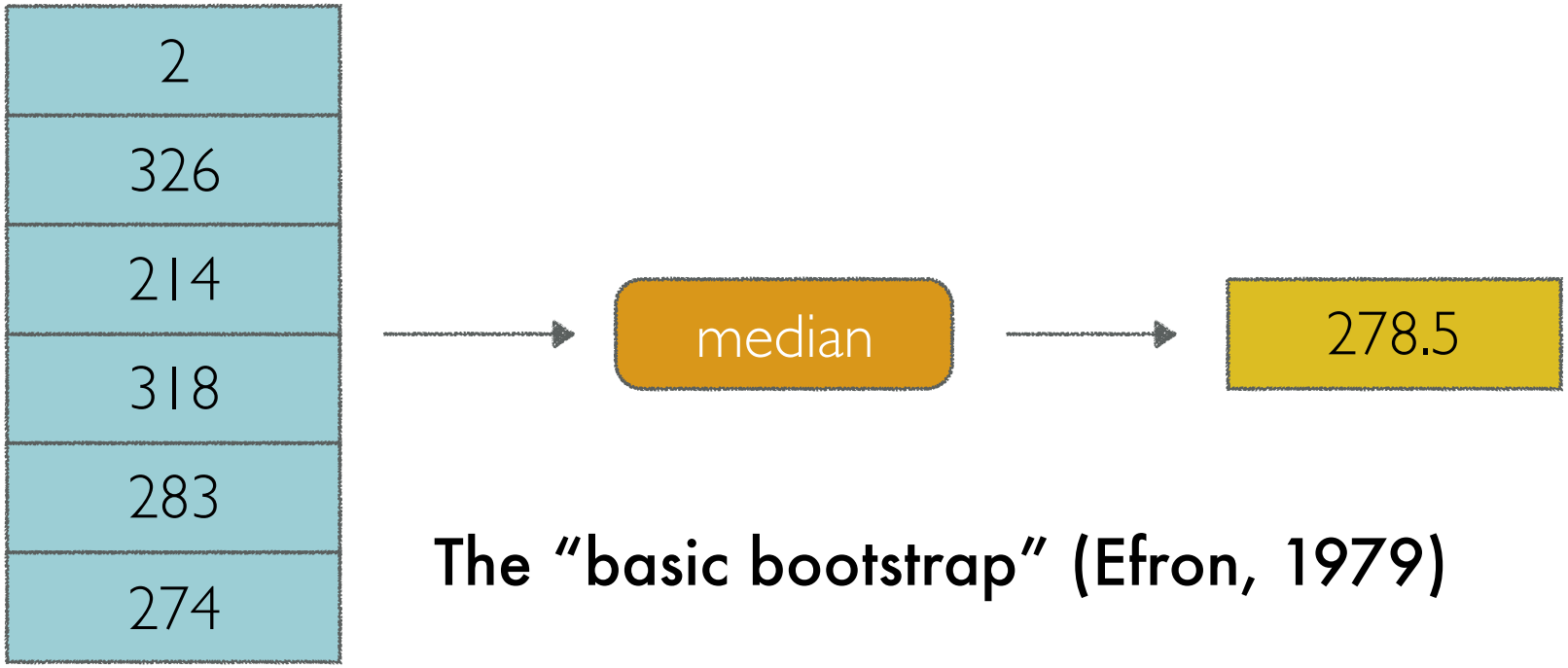


"Donut contains 278.5±5[illegible] kcal."

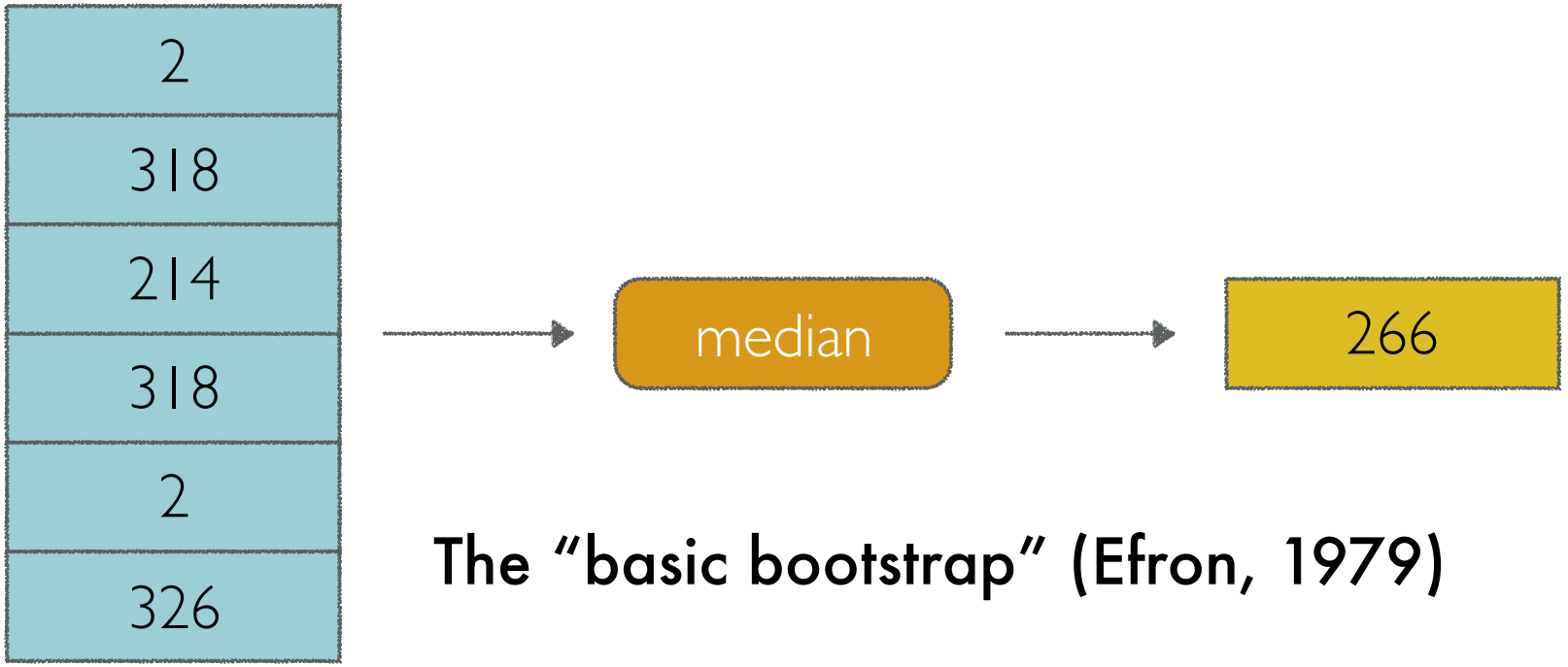
Confidence interval is an unknown function of the distribution , statistic, and sample size.



$$\text{CI for } \bar{x} = \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

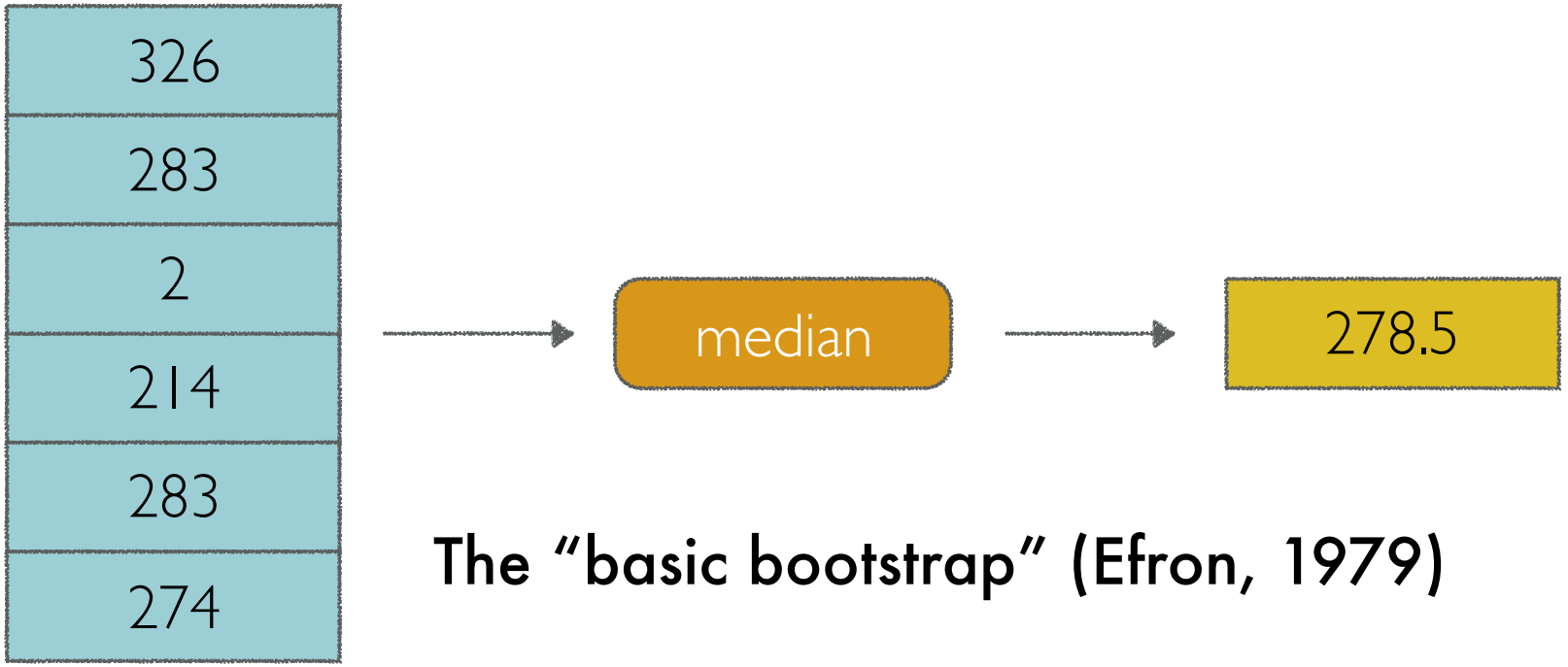


The “basic bootstrap” (Efron, 1979)

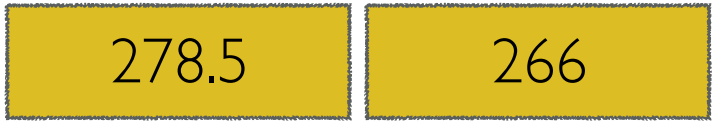


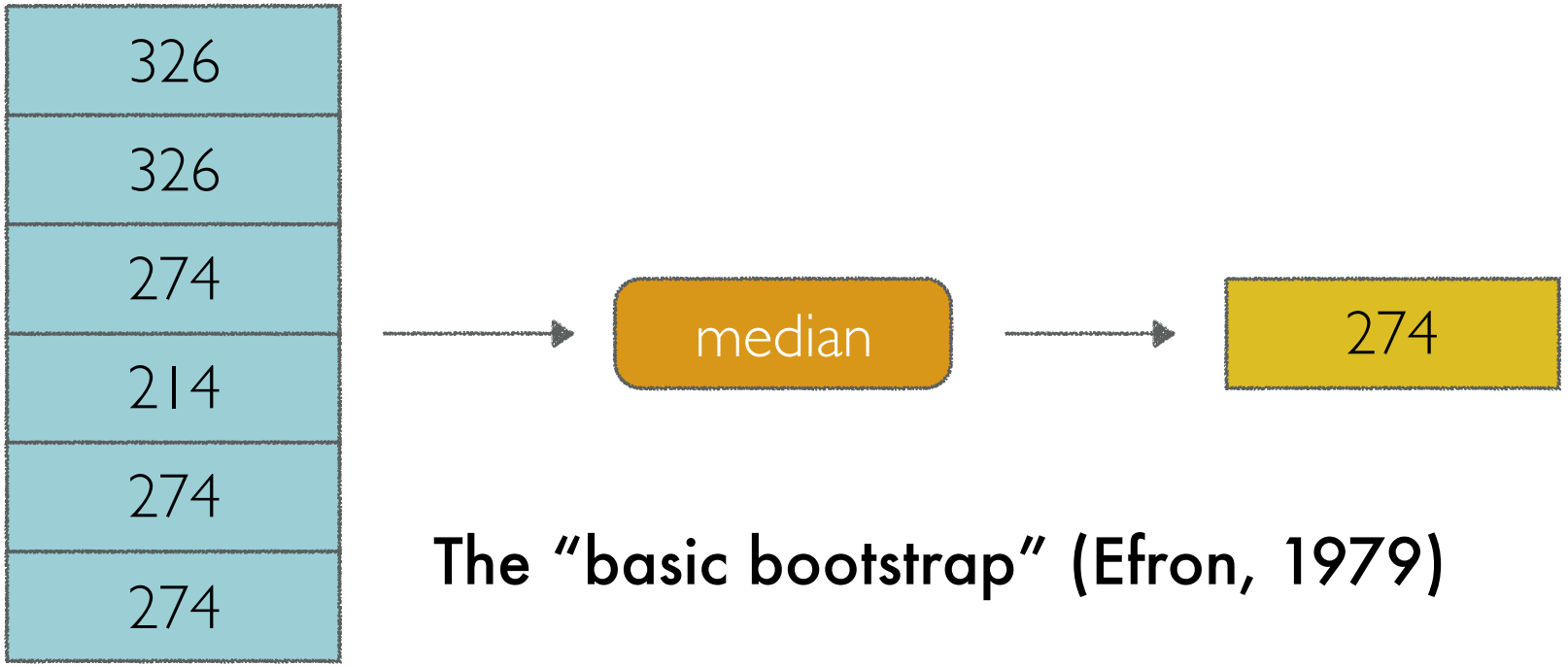
The "basic bootstrap" (Efron, 1979)

278.5



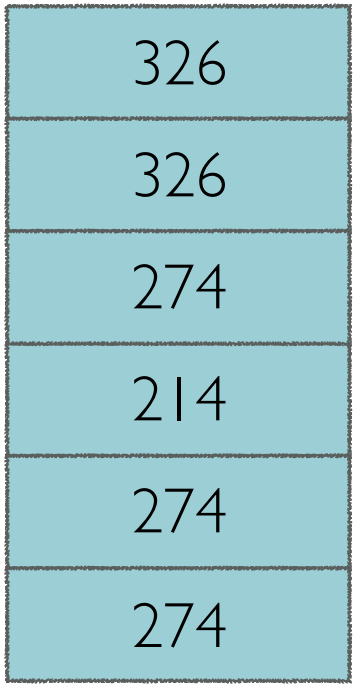
The "basic bootstrap" (Efron, 1979)





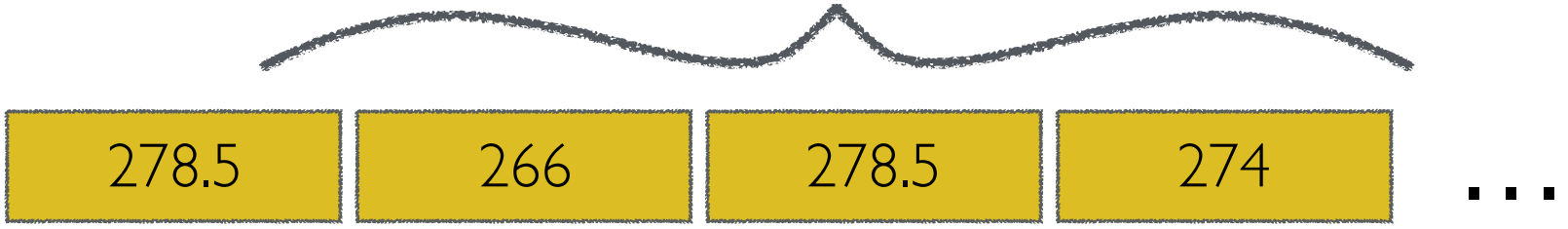
The "basic bootstrap" (Efron, 1979)

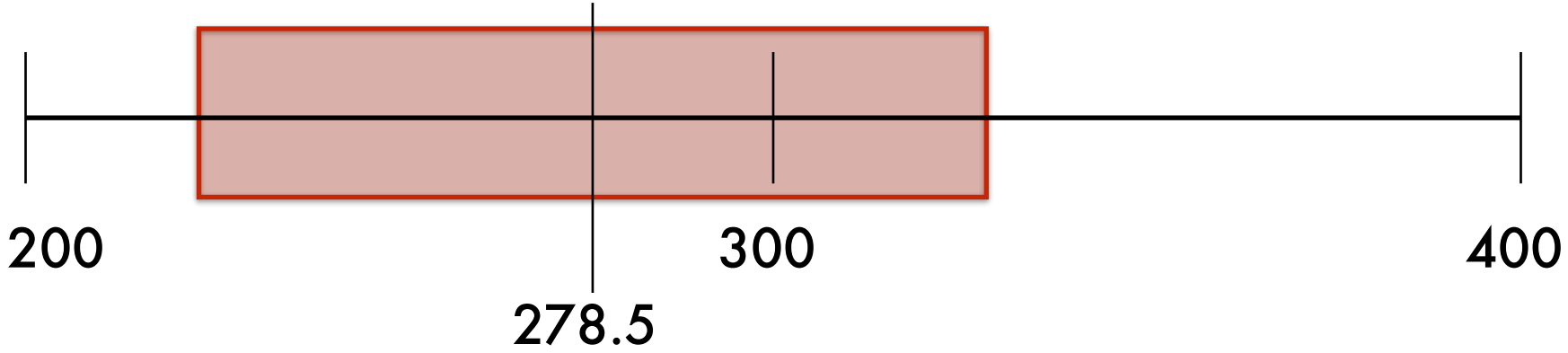
...



Values of median corresponding
to 2.5th and 97.5th percentiles
=> 95% confidence interval

The “basic bootstrap” (Efron, 1979)



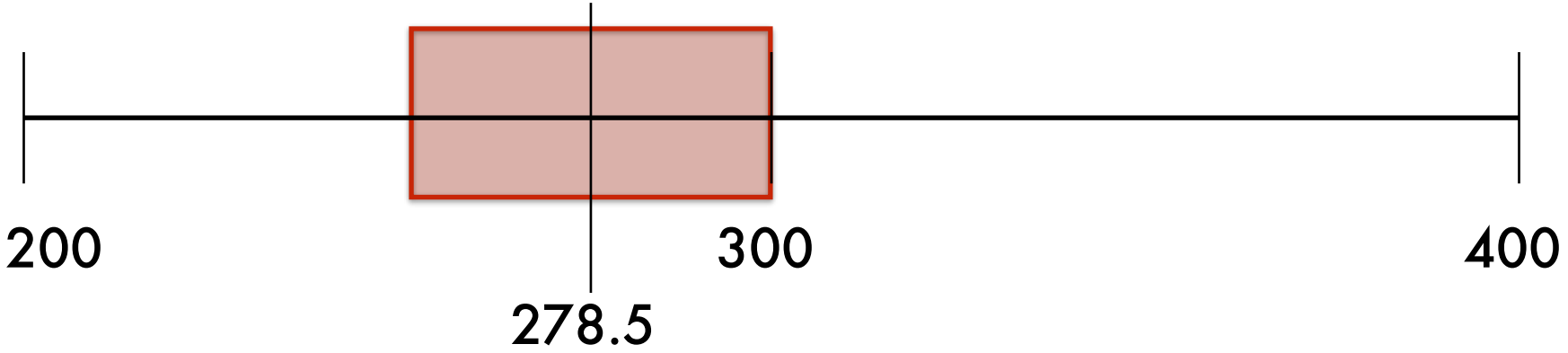


"Donut contains 278.5±50 kcal."

If confidence interval not precise enough, how do we make it tighter?

Confidence interval = function of sample size…

For tighter intervals, ask more people!



"Donut contains 278.5±5.6 kcal."

VoxPL Algorithm

Increase sample size until either

1) interval meets user-defined “tightness” constraint

2) or budget exhausted

```
def numCalories(url: String) = Estimate (
  confidenceInterval = SymmetricCI(50),
  confidence = 0.95,
  budget = 5.00,
  text = "How many calories are
          in the food pictured?",
  imageUrl = url,
  statistic = L1Median
)
```

(mandatory parameters are in red)

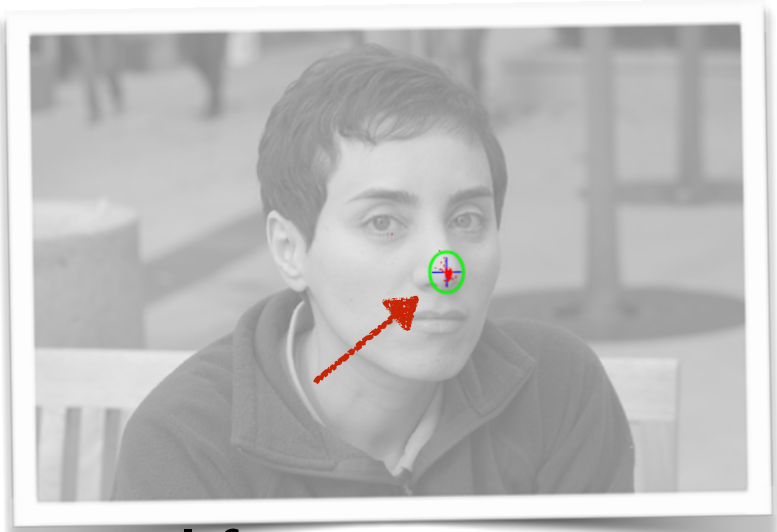
```
val cals = numCalories(breakfast)+
           numCalories(lunch)
```

Computing confidence intervals for *composed* functions: same procedure (because: bootstrap)

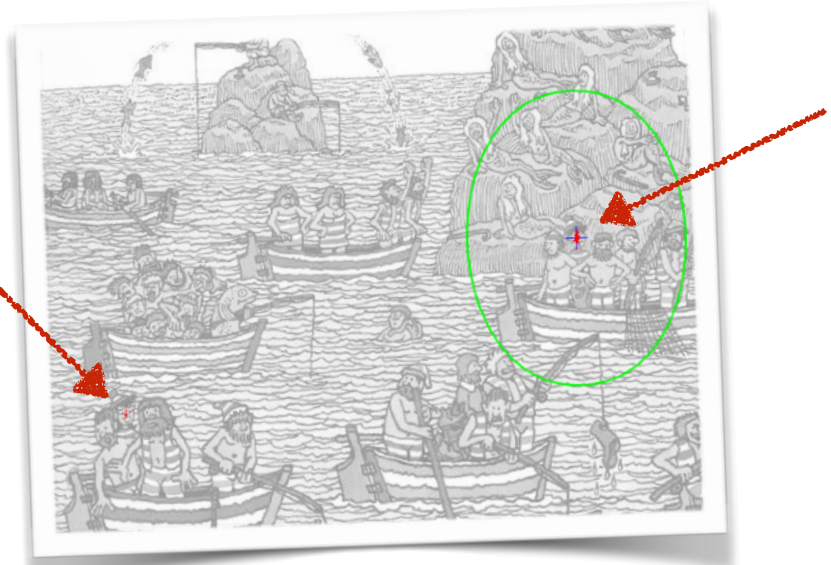
All functions expressible in VoxPL produce valid estimates & empirical confidence intervals

Also ensures that dynamic sample size calculation does not bias results (Bonferroni)

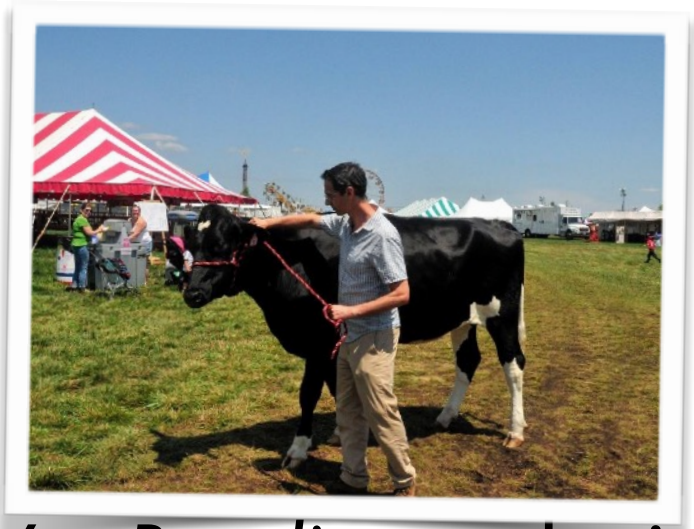
Calorie counting



Facial feature recognition



Difficult visual search

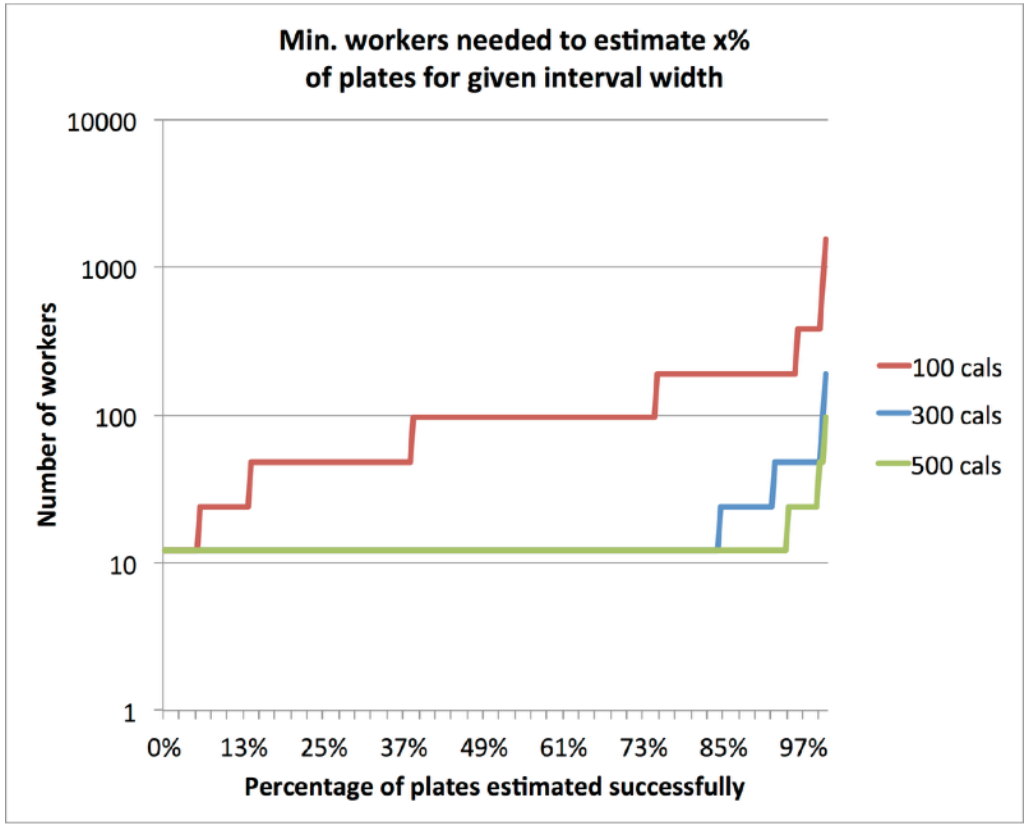


Vox Populi reproduction (NPR Planet Money)

Calorie Counter

- 208 images of school lunches w/ground truth kcal (thanks to Joe Price @ Brigham Young U.)
- IM2Calories MAE: 152.95, SE: 15.61
- VoxPL MAE: 103.08, SE: 6.00

Calorie Counter



Fixed n=200, confidence interval ±41 kcal

Cost: $12 per plate!

VoxPL savings? Depends on constraint

- VoxPL dynamically finds right sample size - can be much cheaper!
  ±250, median cost: $0.32; ±50 cost: $1.28

AutoMan/VoxPL

Automatic budgeting, scheduling, and quality control

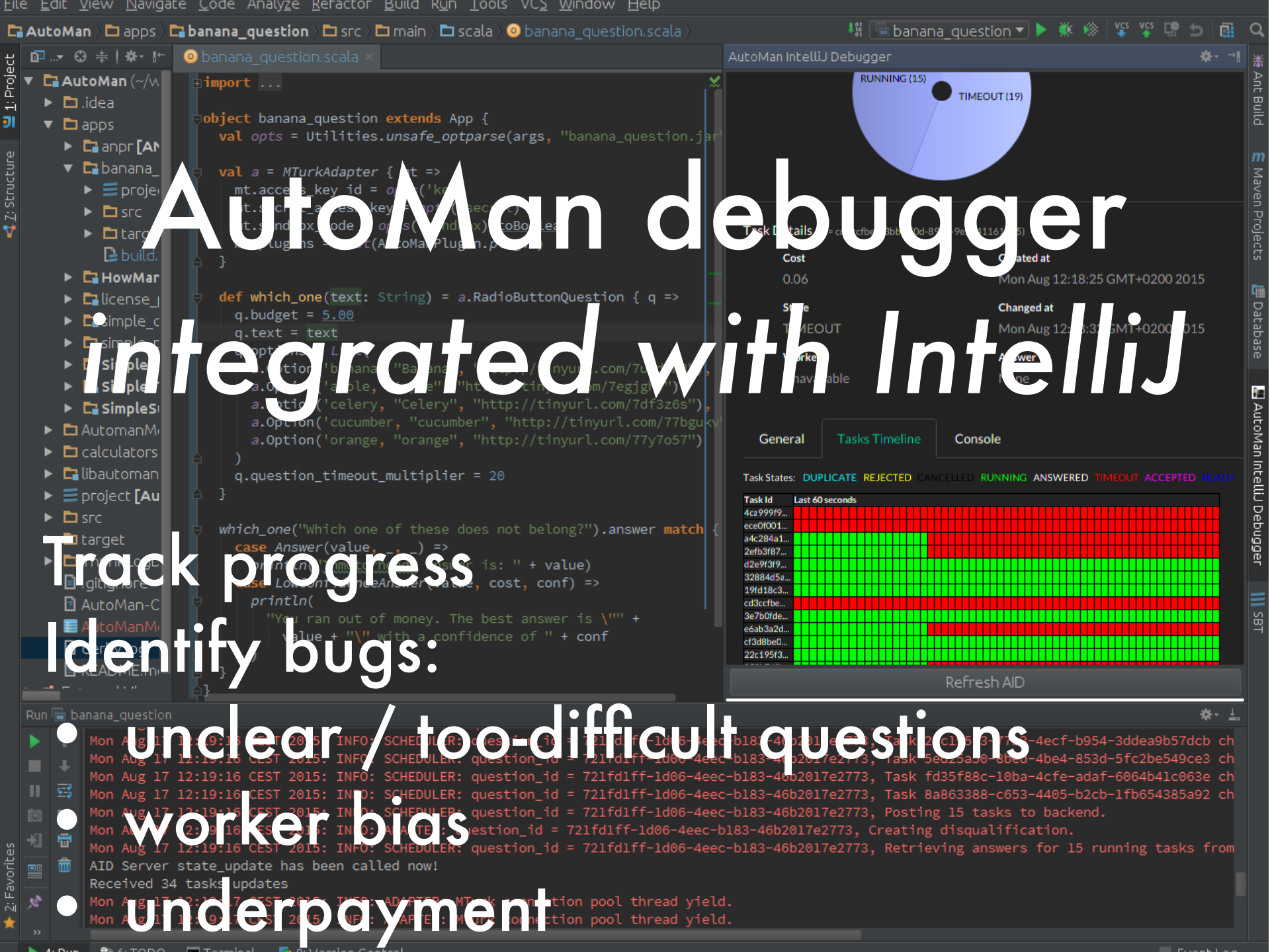
http://github.com/plasma-umass/AutoMan

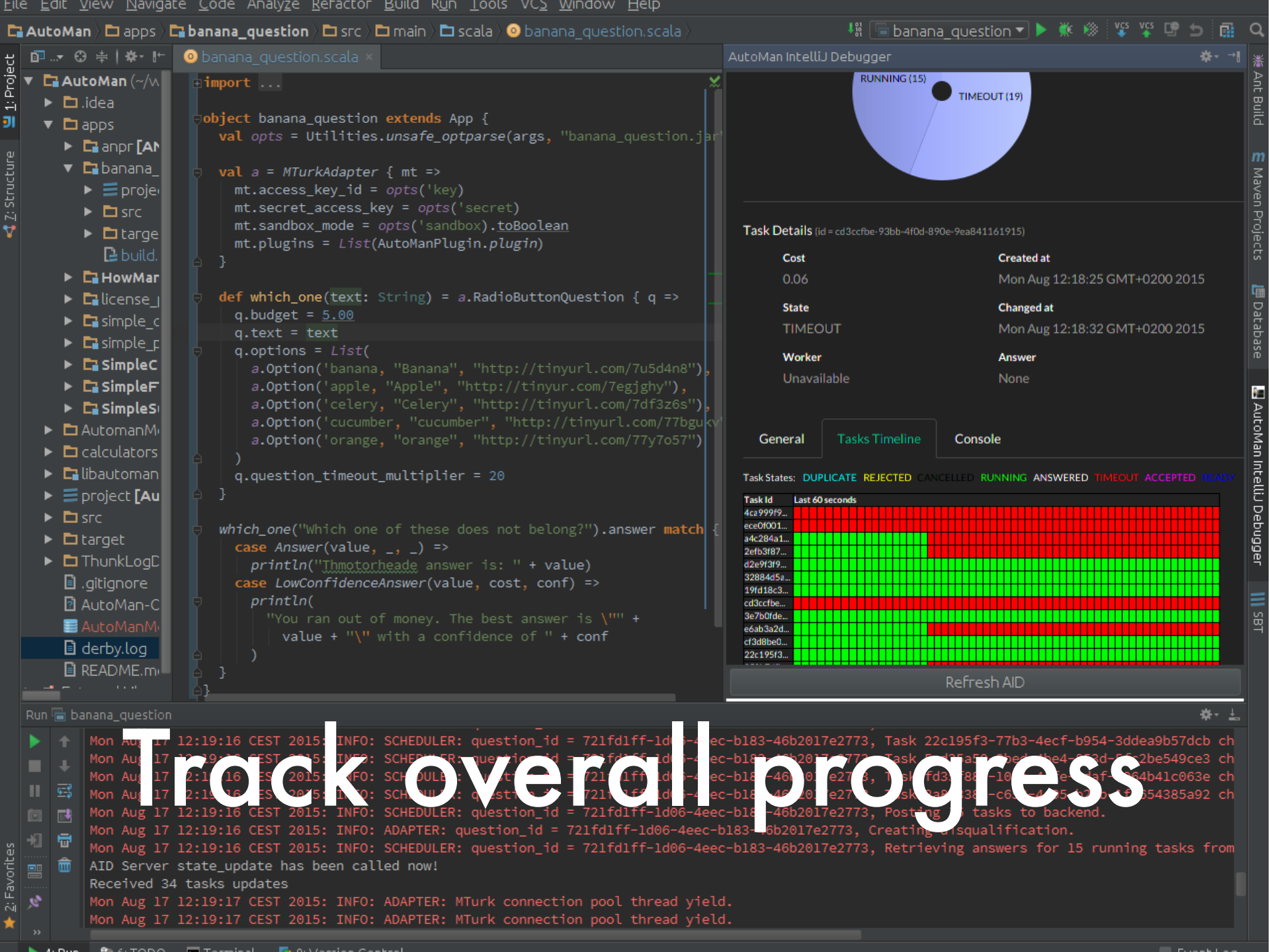
DANIEL BAROWY
CHARLIE CURTSINGER
EMERY BERGER
ANDREW McGREGOR

UMassAmherst

DANIEL GOLDSTEIN
SIDDHARTH SURI
(VoxPL)

Microsoft


AutoMan debugger
integrated with IntelliJ
Track progress
Identify bugs:
• unclear / too-difficult questions
• worker bias
• underpayment


File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
AutoMan apps banana_question src main scala banana_question.scala
banana_question
banana_question.scala
AutoMan (~/w
.idea
apps
anpr [AM
banana_
proje
src
targe
build.
HowMar
license_
simple_c
simple_p
SimpleC
SimpleF
SimpleS
AutomanM
calculators
libautoman
project [Au
src
target
ThunkLogC
.gitignore
AutoMan-C
AutoManM
derby.log
README.m
import ...
object banana_question extends App {
val opts = Utilities.unsafe_optparse(args, "banana_question.jar
val a = MTurkAdapter { mt =>
mt.access_key_id = opts('key)
mt.secret_access_key = opts('secret)
mt.sandbox_mode = opts('sandbox).toBoolean
mt.plugins = List(AutoManPlugin.plugin)
}
def which_one(text: String) = a.RadioButtonQuestion { q =>
q.budget = 5.00
q.text = text
q.options = List(
a.Option('banana, "Banana", "http://tinyurl.com/7u5d4n8"),
a.Option('apple, "Apple", "http://tinyur.com/7egjghy"),
a.Option('celery, "Celery", "http://tinyurl.com/7df3z6s"),
a.Option('cucumber, "cucumber", "http://tinyurl.com/77bguky
a.Option('orange, "orange", "http://tinyurl.com/77y7o57")
)
q.question_timeout_multiplier = 20
}
which_one("Which one of these does not belong?").answer match {
case Answer(value, _, _) =>
println("Thmotorheade answer is: " + value)
case LowConfidenceAnswer(value, cost, conf) =>
println(
"You ran out of money. The best answer is \"" +
value + "\" with a confidence of " + conf
)
}
}
AutoMan IntelliJ Debugger
RUNNING (15)
TIMEOUT (19)
Task Details (id = cd3ccfbe-93bb-4f0d-890e-9ea841161915)
Cost
0.06
Created at
Mon Aug 12:18:25 GMT+0200 2015
State
TIMEOUT
Changed at
Mon Aug 12:18:32 GMT+0200 2015
Worker
Unavailable
Answer
None
General
Tasks Timeline
Console
Task States: DUPLICATE REJECTED CANCELLED RUNNING ANSWERED TIMEOUT ACCEPTED READY
Task Id
Last 60 seconds
4ca999f9...
ece0f001...
a4c284a1...
2efb3f87...
d2e9f3f9...
32884d5a...
19fd18c3...
cd3ccfbe...
3e7b0fde...
e6ab3a2d...
cf3d8be0...
22c195f3...
Refresh AID
Run banana_question
Mon Aug 17 12:19:16 CEST 2015: INFO: SCHEDULER: question_id = 721fd1ff-1d06-4eec-b183-46b2017e2773, Task 22c195f3-77b3-4ecf-b954-3ddea9b57dcb ch
Mon Aug 17 12:19:16 CEST 2015: INFO: SCHEDULER: question_id = 721fd1ff-1d06-4eec-b183-46b2017e2773, Posting 0 tasks to backend.
Mon Aug 17 12:19:16 CEST 2015: INFO: ADAPTER: question_id = 721fd1ff-1d06-4eec-b183-46b2017e2773, Creating disqualification.
Mon Aug 17 12:19:16 CEST 2015: INFO: SCHEDULER: question_id = 721fd1ff-1d06-4eec-b183-46b2017e2773, Retrieving answers for 15 running tasks from
AID Server state_update has been called now!
Received 34 tasks updates
Mon Aug 17 12:19:17 CEST 2015: INFO: ADAPTER: MTurk connection pool thread yield.
Mon Aug 17 12:19:17 CEST 2015: INFO: ADAPTER: MTurk connection pool thread yield.
Track overall progress

extra slides

Which one of these does not belong?


apple


celery


cucumber


orange


banana

Is this a picture of a cowboy?

How many goats?